

STIC-Biotech/ChemLib

From: Chan, Christina
Sent: Tuesday, December 04, 2001 4:01 PM
To: Davis, Minh-Tam; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 09/555342

Please rush. Thanks Chris

-----Original Message-----

From: Davis, Minh-Tam
Sent: Tuesday, December 04, 2001 3:21 PM
To: Chan, Christina
Subject: Rush search request for 09/555342

Please search in commercial database and in issued patent files:

- 1) SEQ ID NOS: 1 and 2
- 2) Amino acids 1-374 of SEQ ID NO:2.
- 3) Amino acids 544-737 of SEQ ID NO:2
- 4) Amino acids 764 to 854 of SEQ ID NO:2.
- 5) Nucleotides 49-3183 of SEQ ID NO:1.
- 6) Oligomer search for at least more than 90 amino acids of SEQ ID NO:2

Thank you,
MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

priority date:

Nov 97

CRFF

if Contact:
Sheppard

Searcher: _____
Phone: tel: 308-4499
Location: _____
Date Picked Up: _____
Date Completed: 12/7/01
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw.model

Run on: December 6, 2001, 08:46:08 ; Search time 72.4 Seconds
(without alignments)
1069.152 Million cell updates/sec

Title: US-09-555-342A-2
Perfect score: 5463
Sequence: 1 MGEIQRPTPGSLGAPENS.....SATSSASRPHVLSHKESLVY 1045

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.ll01.*

1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	5463	100.0	1045	20 AAY07482	Human chondrocyte-
2	5459	99.9	1045	21 AAY91947	Human cytoskeleton
3	1737	31.8	395	21 AAB54227	Human pancreatic c
4	720	13.2	1005	21 AAY91946	Human cytoskeleton
5	626	11.5	687	22 AAB95258	Human protein sequ
6	541	9.9	913	18 AAW12522	Protein tyrosine p
7	541	9.9	913	20 AAY25156	Human ptpH1 protei
8	520.5	9.5	766	21 AAY51248	Rat actin-binding
9	419	7.7	369	22 AAU04486	Human protein tyro
10	409	7.5	432	22 AAB93568	Human protein sequ
11	396	7.2	1174	19 AAW67438	Human protein tyro

12	396	7.2	1174	20 AAY34158	Human protein tyro
13	367.5	6.7	635	21 AAB53356	Human colon cancer
14	356.5	6.5	586	20 AAY27443	Amino acid sequenc
15	342.5	6.3	305	22 AAM25892	Human protein sequ
16	335	6.1	590	20 AAW94458	Human neurofibroma
17	335	6.1	591	18 AAW09648	Human merlin prote
18	335	6.1	595	15 AAR60398	Merlin protein enc
19	335	6.1	596	20 AAW94459	Human neurofibroma
20	334	6.1	584	20 AAW94457	Mouse neurofibroma
21	334	6.1	591	20 AAW94456	Mouse neurofibroma
22	334	6.1	596	20 AAW94455	Mouse neurofibroma
23	333	6.1	495	22 AAM41770	Human polypeptide
24	332	6.1	584	18 AAW09647	Mouse merlin prote
25	330	6.0	591	18 AAW09646	Mouse merlin prote
26	330	6.0	596	18 AAW09645	Mouse merlin prote
27	320.5	5.9	445	22 AAM39984	Human polypeptide
28	308	5.6	2485	21 AAB19343	Amino acid sequenc
29	306	5.6	279	22 AAM25733	Human protein sequ
30	300.5	5.5	2466	16 AAR71498	Human protein tyro
31	300.5	5.5	2466	19 AAW75999	Intracellular prot
32	300.5	5.5	2466	21 AAY90272	Human ptpH1 phosph
33	287.5	5.3	450	22 AAB92607	Human protein sequ
34	266	4.9	1105	22 AAM25567	Human protein sequ
35	264.5	4.8	2861	18 AAW27227	Human TRIO phospho
36	256.5	4.7	619	22 AAB97025	Human colon carcin
37	245	4.5	142	22 AAG74414	Human colon cancer
38	244.5	4.5	1715	21 AAY57449	Mouse Ees1L protei
39	242	4.4	1520	20 AAY41010	Amino acid sequenc
40	240.5	4.4	1527	22 AAU01184	Rat glutamate tran
41	239	4.4	1683	21 AAY71160	Rat phosphodiester
42	235.5	4.3	523	22 AAM39338	Human polypeptide
43	234.5	4.3	647	22 AAM41124	Human polypeptide
44	227.5	4.2	580	22 AAW81349	Human guanine nucl
45	223	4.1	1658	21 AAY57450	Mouse Ees2L protei

ALIGNMENTS

RESULT 1
AAY07482
ID AAY07482 standard; Protein; 1045 AA.
XX AC AAY07482;
XX DT 17-AUG-1999 (first entry)
XX DE Human chondrocyte-derived protein CDEP.
XX KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
XX KW Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;
XX KW drug.
XX OS Homo sapiens.
XX PN WO9928458-A1.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98WO-JP05348.
XX PR 27-NOV-1997; 97JP-0342060.
XX (CHUS) CHUGAI SEIYAKU KK.
XX PI Kato Y, Kawamoto T, Koyano Y;
XX DR WPI; 1999-371117/31.
XX DR N-PSDB; AAX79183.
XX PT Protein CDEP expressed in differentiated chondrocytes, and gene
XX encoding it

PS Claim 2; Fig 1; 59pp; Japanese.

XX This sequence represents a protein (CDEP) expressed in differentiated
CC human foetal chondrocytes, which contains an erin-like domain, a Db1
CC homology (DH) domain and a pleckstrin homology (PH) domain. The encoding
CC nucleic acid or protein can be used in the investigation and treatment of
CC cancers and arthritic diseases (including chronic rheumatoid arthritis),
CC or for screening of candidate anticancer drugs.

XX Sequence 1045 AA;

Query Match 100.0%; Score 5463; DB 20; Length 1045;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1045; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGEIEORPTPGSLGAPENSGISTLERGQKPPPTPSGKLVSIKIQMLDDTQAFQVQRA 60
Db 1 mgeiedrptpgslgapensgistlergqkppptpsgklvsiqimldtqafevpgra 60
QY 61 PGKVLDAVCNHLNLVEGDYFGLERPDHKKITVWLDLLKPIVKQIRRPKHVVVVFVVKFF 120
Db 61 pgkvlldavcnhlnlvegdylfglefepdhkkitvwlldllkpivkqirrpkhvvvkvvkkff 120
QY 121 PPDHTQLOELTRYLPALQVQDLAQRCLTNDTSAAALLSHIVQSEIGDFDEALDREHL 180
Db 121 ppdhtqlqeeltrylfalqvqdlagrltncndtsaallshivqseigdfdealdrehl 180
QY 181 AKNKYIPOQDALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKI 240
Db 181 aknkyipqddaledkivefhnnhigotpaesdfolleiarremygirllhpakdregtki 240
QY 241 NLAIVANTGILVFGFTKINAFNAWVKLSFKRKRFLIKLRPDANSAYODTLEFLMASRD 300
Db 241 nlaivantgilvfgftkinafnawvklsfkrkrflilkrpdansayodtleflmasrd 300
QY 301 FCKSEWKICVEHHAFFLPEEPKPKPVLFSGSGSFRFSGRTQKQVLDYVKEGGHKYQ 360
Db 301 fcksewkicvehhaflpeepkpkpvlfsgsgsfrfsgrtqkvldyvkeggkkyq 360
QY 361 FERKSHKHSIRSLASQPTLSEVLEQSQQSTLSFTFGGASPGSQSCRRGKEPKVSAG 420
Db 361 ferkshkhsirslasqptlsevleqsgqstlsftfgegaspsgsgsrrgkepkvsag 420
QY 421 EPQSHSPAPRSPAGNKQADGAASAPTEEEVEVWKDRTOQSKPQPPQSTGSLTGSPLH 480
Db 421 epqshspaprrspagnkqadgaasapteeevewkdrtoqskpqpstgstsltgsplh 480
QY 481 SELSVNSQGGVAPANTVLTSPNLTSPDKQASPLISPLINDOACPRDDEDEGRKRFPDOK 540
Db 481 selsvnsqggvapantvltspnltspdkqasplisplindqacprtdedegrkrfptdk 540
QY 541 AYFIAKEVSTERTYIKLDEVITSWFQSVSKEDAMPEALKSLIPNFPPLHKFHTNFKL 600
Db 541 ayfiakvestertyikldevitswfskvedampealkslipnfpplhkfhntnflk 600
QY 601 ETEORLALWEGRSNAQIRDYQRIGDVMKLNIOCMKHLAHLWKHSALALENGIKSSRR 660
Db 601 eteoralwegrснаqirdyqrigdvmklniocmkhlaahlwkhsalealengikssrr 660
QY 661 LENFCRDFELQVCYLPLNTFLLRPLRLHMYKQVLERLCKKHPPSHADFRDCRAALAEI 720
Db 661 lenferdfelqvcylplntfllrplrlhmykqvlerlckkhppshadfrdcraalaei 720
QY 721 TEMVAQLHGTMIKMFQKHLKDLIGIDNLVPGREFIRLGSLSKLSGKGLQORMFF 780
Db 721 temvaqlhgtmikmfqkhlkdligidnlvpgrefirlgslslsklsgkglqormff 780
QY 781 LFNVDVLLYTSRGLTASNQKFKVHQLPLYGMTIESEDEWGVPHCLTLRGOROSIIVAAS 840
Db 781 lfnvdvlllytsrgltasnqkvhqlplygmtieesedewgvphcltlrggrsiivaas 840
QY 841 RSEMEKWVEDIOWAIDLAEKSSPAPFLASSPPDNKSPDEATAADQESDEDDLSASRTSL 900

Db 841 rsemekwvedigwaidlaeksspapflassppdnkspdeataadqesddlsasrtsl 900
QY 901 ERQAPHRGNTMWHVCHRTSYSMVDFSTAVENQLSGNLLRKFKNNGWOKLWVFTNFC 960
Db 901 erqaphrgntmwhvchrtssvmvdfstavenqlsgnllrxfknsngwqklwvftnfc 960
QY 961 LFFYKSHQDNHPLASLPLGYSLTIPSESENQKQVFKLHFKSHVYVYFRASEYTFERW 1020
Db 961 lffykshqdnhplaslpligysltipseseniqkvfklhfkshvyvyyfraeseytfew 1020
QY 1021 MEVIRSATSSASRPHVLSHESLVY 1045
Db 1021 mevirsatssasrphvlskheslvy 1045
RESULT 2
AA91947
ID AA91947 standard; Protein; 1045 AA.
XX
AC AA91947;
XX
DT 19-JUL-2000 (first entry)
XX
DE Human cytoskeleton associated protein 2 (CYSKP-2).
XX
KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 23 /note= "potential phosphorylation site"
FT Modified-site 36 /note= "potential phosphorylation site"
FT Modified-site 41 /note= "potential phosphorylation site"
FT Domain 47..85 /note= "signature sequence"
FT Modified-site 92 /note= "potential phosphorylation site"
FT Domain 94..123 /note= "signature sequence"
FT Domain 144..190 /note= "signature sequence"
FT Modified-site 150 /note= "potential phosphorylation site"
FT Modified-site 152 /note= "potential N-glycosylation site"
FT Domain 196..249 /note= "signature sequence"
FT Modified-site 207 /note= "potential phosphorylation site"
FT Domain 261..279 /note= "signature sequence"
FT Modified-site 270 /note= "potential phosphorylation site"
FT Modified-site 336 /note= "potential phosphorylation site"
FT Modified-site 340 /note= "potential phosphorylation site"
FT Modified-site 343 /note= "potential phosphorylation site"
FT Modified-site 366 /note= "potential phosphorylation site"
FT Modified-site 370 /note= "potential phosphorylation site"
FT Modified-site 396 /note= "potential phosphorylation site"
FT Modified-site 408 /note= "potential phosphorylation site"

FT Modified-site 418 /note= "potential phosphorylation site"
 FT Modified-site 448 /note= "potential phosphorylation site"
 FT Modified-site 495 /note= "potential N-glycosylation site"
 FT Modified-site 525 /note= "potential phosphorylation site"
 FT Modified-site 538 /note= "potential phosphorylation site"
 FT Modified-site 542 /note= "potential phosphorylation site"
 FT Modified-site 549 /note= "potential phosphorylation site"
 FT Modified-site 551 /note= "potential phosphorylation site"
 FT Modified-site 571 /note= "potential phosphorylation site"
 FT Modified-site 657 /note= "potential phosphorylation site"
 FT Modified-site 658 /note= "potential phosphorylation site"
 FT Modified-site 706 /note= "potential phosphorylation site"
 FT Modified-site 770 /note= "potential phosphorylation site"
 FT Modified-site 789 /note= "potential phosphorylation site"
 FT Modified-site 811 /note= "potential phosphorylation site"
 FT Modified-site 815 /note= "potential phosphorylation site"
 FT Modified-site 826 /note= "potential phosphorylation site"
 FT Modified-site 839 /note= "potential phosphorylation site"
 FT Domain 770..773 /note= "signature sequence"
 FT Modified-site 840 /note= "potential phosphorylation site"
 FT Modified-site 842 /note= "potential phosphorylation site"
 FT Modified-site 872 /note= "potential phosphorylation site"
 FT Modified-site 878 /note= "potential phosphorylation site"
 FT Modified-site 883 /note= "potential phosphorylation site"
 FT Modified-site 889 /note= "potential phosphorylation site"
 FT Modified-site 898 /note= "potential phosphorylation site"
 FT Modified-site 919 /note= "potential N-glycosylation site"
 FT Modified-site 923 /note= "potential phosphorylation site"
 FT Modified-site 966 /note= "potential phosphorylation site"
 XX WO200017355-A2.
 XX 30-MAR-2000.
 XX 17-SEP-1999; 99WO-US21565.
 XX 18-SEP-1998; 98US-0172226.
 XX 27-APR-1999; 99US-0131321.
 XX (INCY-) INCYTE PHARM INC.
 XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
 PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
 XX

DR WPI; 2000-283582/24.
 DR N-PSDB; AAA08582.
 XX Human cytoskeleton associated proteins, used to treat cell
 PT proliferative, autoimmune/inflammatory, vesicle trafficking,
 PT neurological, cell motility, reproductive and muscle disorders
 XX Claim 1; Page 82-84; 113pp; English.
 PS
 XX AAY91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1
 CC to CYSKP-16) respectively. The sequences can be used to treat and
 CC diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle
 CC trafficking, neurological, cardiovascular, cell motility, reproductive
 CC and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to
 CC CYSKP-16 can be used to treat or prevent disorders associated with
 CC decreased expression or activity of CYSKP (claimed), for example,
 CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
 CC cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,
 CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
 CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
 CC and trauma. CYSKP antagonists can be used to treat or prevent a
 CC disorder associated with increased expression or activity of CYSKP
 CC (claimed).
 XX Sequence 1045 AA;
 SQ
 Query Match 99.9%; Score 5459; DB 21; Length 1045;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MGEIQRTPTGSRGAPENSIGSTLERGQKPPPTSGKLVSTIKIOMLDTQAFEPQRA 60
 DB 1 mgeieqrptpgsrlgapensigstlergqkppptsgklvstikimldtqafevpgqa 60
 QY 61 PGKVLDAVCNHLNVEGDYFGLFEPDHHKITVMDLLKPIVKQIRRRPKHVVFVKPF 120
 DB 61 pgkvlldavcnhlnvegdylfglefpdhkkitvmdllkpivkqirrrpkhvvvkvkff 120
 QY 121 PPDHTQLQBELTRYLFALQVKQDLAQGRUTCNDTSAALISHIVQSEIGDFDEALDREHL 180
 DB 121 ppdhtqlqeeltrylfalqvkdlaqgrlucndtsaallishivqseigdfdealdrehl 180
 QY 181 ANKYIPOODALEDKIVEFHNNHIGOTPAESDFOLLEIARRLMEYGIIRLHPAKDREGTKI 240
 DB 181 ankypipqgdaldedkivefhnnhigotpaesdfqlleiarriemylgrrlhpakdregtki 240
 QY 241 NLAVANTGLVFGQETKINAFNWKVKLSFKRKFLIKLRPDANSAYODTLEFLMASRD 300
 DB 241 nlavantgilvfggftkinafnwkvrlsfkrkffliklrpdansayodtleflmasrd 300
 QY 301 FCKSFWKICVEHAFRFLPEPKPKPVLFSGSGSFRSGRTQKQVLDYVKEGGHKVQ 360
 DB 301 fcksfwkicvehahfrifeepkpkpvlfsrgsgsfrsgrtqkqvldyvkeggkhkvq 360
 QY 361 FERKHSKIHISRLASQPTLNSEVLEQSQSTSTFGEAESPGQSCRRKQKPKVSAG 420
 DB 361 ferkhskihsrlasqptelnsevlseqsststfgeaespggqscrrkqkpkvsgag 420
 QY 421 EPGSHPSAPRRSPAGNKQADGAASAPTEEEVVKDRTQOSKPOPPQSTGSLTGSPLH 480
 DB 421 epgshpsaprrspagnkqadgaasapteeeevvkdrtooskpopppqstgsltgsphl 480
 QY 481 SELSVNSGGVAPANVTLSPLNSPDTKQASPLISPLLNDAQPRDDEDEGRKRKRPPTDK 540
 DB 481 selsvnsggvapanvtlspnlspdtkqasplispllndaqprtdedegrkrkrfptdk 540
 QY 541 AYFIAKEVSTTERTYKLDLEVITSWFQSTVSKEDAMPEALKSLIFPNFPLHKHFTNFK 600
 DB 541 ayfiakevsttertykldlevitswfqstvsikedampealkslifpnfplhkhtnflk 600
 QY 601 ETEQRIALWEGRSNAQIRDYQIRGIDVWMLKNIQGMKHLAAHLWKHSALEALENGKSSRR 660
 DB 601 eteqrialwegrsnaqirdyqirgidvwmklniqgmkhlaahlwkhsalealengkssrr 660

Db 601 eiegrlalwegrnagirdyqrigdvmlknigmkhlaahlwkhsealealengikssrr 660
QY 661 LENFCRDFELQKCYLPLNTFLLRLHMLHYKQVLERLCKHPPSHADPRDCRAALAEI 720
Db 661 lenfcdrfclqkvcyplntfllrplhrlmhykqvlerlckhppshadfrdraael 720
QY 721 TEMVAQLHGTMIRKMFQKHLKDLIGDNIWVPGREFIRGLSKSLSGKLGQORMFF 780
Db 721 temvaqlhgtmkmfqkhlkdlidnvwpgrefirglsgslsgkqlgrmff 780
QY 781 LFNVDLLYTSRGTLTASNQFVHGOLPLVGMTESEDEWGVPHCLTLRGQRQIIVAASS 840
Db 781 lfnvdllytsrgtltasnfqvhgolgplvgtmteesedewgvphcltlrgqrqiivaass 840
QY 841 RSEMEKWVEDIQWIDLAERSSSPAPFLASSPPDNKSPDEATAADQESDDLSASRTSL 900
Db 841 rsemekwvediqwidlaerssspapflassppdnkspdeataadqesddlsasrtsl 900
QY 901 ERQAPHRGNTVMVHVRNRTSVSMVDFSTAVENQLSGNLLRKFKNSGMOKLWVFTNFC 960
Db 901 erqaphrgntvmvvhvnrntsvsmvdfstavenqlsgnllrkfknsngmklwvvtncf 960
QY 961 LFPYKSHQDNHPLASPLLYGSLTIPSESENIQKDYVFKLHFKSHVYVYFRAESEYTFERW 1020
Db 961 lfykshqdnhplaslpllygsltipseesenikdyvfkhlkshvyvvyfraeseytferw 1020
QY 1021 MEVIRSATSSASRPHVLSHKSLEY 1045
Db 1021 mevirsatssasrphvlskley 1045

RESULT 3
AAB54227
ID AAB54227 standard; Protein; 395 AA.
XX AAB54227;
AC AAB54227;
XX
DT 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:679.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
OS Homo sapiens.
XX
XX WO200055320-A1.
PN
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
XX WPI; 2000-579444/54.
DR N-PSDB; AAC98992.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition
XX
XX Claim 11; Page 1115-1116; 1379pp; English.
PS
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated

CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

Sequence 395 AA;

Query Match 31.8%; Score 1737; DB 21; Length 395;
Best Local Similarity 97.7%; Pred. No. 1.3e-135;
Matches 335; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 565 WFGSTVSKEDAMPEALKSLIFPNFELPHKFTNLFKEIORLALWEGRSNAQIRDYORIG 624
Db 21 wfgstvsksedampealkslifpnfepfhkftnlkfeiorlalgwgrsnagirdyqrig 80
QY 625 DYMLKNIQCMKHLAHLWKHSEALELENGIKSSRLENFCRDFELQKCYLPLNTFLLR 684
Db 81 dymlkniqcmkhlahlwkhsalealeengikssrlefcdrfclqkvcyplntfllr 140
QY 685 PLHRLMHYQVLERLCKHHPPSHADPRDCRAALAEITEMVAQLHGTMIRKMFQKHLK 744
Db 141 plhrlmhyqvlerlckhphshadfrdcraalaeitemvaqlhgtmkmfqkhlk 200
QY 745 KDLIGDNIWVPGREFIRGLSKSLSGKLGQORMFFLNDVLLYTSRGTLTASNQFVHGQ 804
Db 201 kdligidnvwpgrefirglsgslsgkqlgrmfflndvlytsrgtltasnfqvhqg 260
QY 805 LPLVGMTESEDEWGVPHCLTLRGQRQIIVAASSRSEMEKWVEDIQWIDLAERSSSP 864
Db 261 lplvgmtteesedewgvphcltlrgqrqiivaassrsemekwvediqwidlaersssp 320
QY 865 APEFLASSPPDNKSPDEATAADQESDDLSASRTSLERQAPHR 907
Db 321 apeflassppdnkspdeataadqesddlsas-----phr 355

RESULT 4
AAY91946
ID AAY91946 standard; Protein; 1005 AA.
XX
AC AAY91946;
XX
DT 19-JUL-2000 (first entry)
DE Human cytoskeleton associated protein 1 (CYSKP-1).
XX
XX Cytoskeleton associated protein; CYSKP-1; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Modified-site 7 /note= "potential phosphorylation site"
FT

FT Modified-site 15 /note= "potential phosphorylation site"
 FT Modified-site 36 /note= "potential glycosylation site"
 FT Modified-site 39 /note= "potential phosphorylation site"
 FT Modified-site 58 /note= "potential phosphorylation site"
 FT Modified-site 151 /note= "potential phosphorylation site"
 FT Modified-site 159 /note= "potential phosphorylation site"
 FT Modified-site 180 /note= "potential phosphorylation site"
 FT Modified-site 194 /note= "potential phosphorylation site"
 FT Modified-site 208 /note= "potential phosphorylation site"
 FT Modified-site 212 /note= "potential phosphorylation site"
 FT Modified-site 223 /note= "potential phosphorylation site"
 FT Modified-site 225..263 /note= "signature sequence"
 FT Domain 272..300 /note= "signature sequence"
 FT Domain 321..367 /note= "signature sequence"
 FT Modified-site 323 /note= "signature sequence"
 FT Modified-site 379..408 /note= "potential phosphorylation site"
 FT Domain 381 /note= "signature sequence"
 FT Modified-site 402 /note= "potential phosphorylation site"
 FT Modified-site 440..458 /note= "signature sequence"
 FT Domain 449 /note= "signature sequence"
 FT Modified-site 518 /note= "potential phosphorylation site"
 FT Modified-site 530 /note= "potential phosphorylation site"
 FT Modified-site 543 /note= "potential phosphorylation site"
 FT Modified-site 544 /note= "potential phosphorylation site"
 FT Modified-site 548 /note= "potential phosphorylation site"
 FT Modified-site 614 /note= "potential phosphorylation site"
 FT Modified-site 623 /note= "potential phosphorylation site"
 FT Modified-site 647 /note= "potential phosphorylation site"
 FT Modified-site 658 /note= "potential phosphorylation site"
 FT Modified-site 673 /note= "potential phosphorylation site"
 FT Modified-site 682 /note= "potential phosphorylation site"
 FT Domain 718..721 /note= "signature sequence"
 FT Modified-site 730 /note= "potential phosphorylation site"
 FT Modified-site 744 /note= "potential phosphorylation site"
 FT Modified-site 746 /note= "potential phosphorylation site"
 FT Modified-site 748 /note= "potential phosphorylation site"
 FT Modified-site 766 /note= "potential phosphorylation site"

FT Modified-site /note= "potential phosphorylation site"
 FT Modified-site 828 /note= "potential phosphorylation site"
 FT Modified-site 854 /note= "potential phosphorylation site"
 FT Modified-site 879 /note= "potential phosphorylation site"
 FT Modified-site 884 /note= "potential phosphorylation site"
 FT Modified-site 890 /note= "potential phosphorylation site"
 FT Modified-site 944 /note= "potential phosphorylation site"
 FT Modified-site 952 /note= "potential phosphorylation site"
 FT Modified-site /note= "potential phosphorylation site"
 XX WO200017355-A2.
 PN 30-MAR-2000.
 PD
 XX
 XX 17-SEP-1999; 99WO-US21565.
 XX
 PR 18-SEP-1998; 98US-0172226.
 PR 27-APR-1999; 99US-0131321.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
 PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
 XX
 DR WPI; 2000-283582/24.
 DR N-PSDB; AAA08581.
 XX
 PT Human cytoskeleton associated proteins, used to treat cell
 PT proliferative, autoimmune/inflammatory, vesicle trafficking,
 PT neurological, cell motility, reproductive and muscle disorders
 XX
 PS Disclosure; Page 79-82; 113pp; English.
 XX
 CC AAY91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1
 CC to CYSKP-16) respectively. The sequences can be used to treat and
 CC diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle
 CC trafficking, neurological, cardiovascular, cell motility, reproductive
 CC and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to
 CC CYSKP-16 can be used to treat or prevent disorders associated with
 CC decreased expression or activity of CYSKP (claimed), for example,
 CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
 CC cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,
 CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
 CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
 CC and trauma. CYSKP antagonists can be used to treat or prevent a
 CC disorder associated with increased expression or activity of CYSKP
 CC (claimed).
 XX
 SQ Sequence 1005 AA;

Query Match 13.2%; Score 720; DB 21; Length 1005;
 Best Local Similarity 31.2%; Pred. No. 1-7e-50;
 Matches 192; Conservative 102; Mismatches 222; Indels 100; Gaps 18;

QY 34 TPGKGLVSIKIQMLDDTQAEFVQPAPGVKVLDDAVNCNHLNLVEGDFGLFPPDHKKITV 93
 Db 212 tkktvtqckvtllldgteyscdlekhakgvlfkdcvchlnllekyflllfgespeqn 271
 QY 94 WLDLLKPIVKQIRPKKVVVKKVFFPPDPHTQLOELTVLFAOVKQDLAOGRLTCND 153
 Db 272 wldpakeikrqlnlpwift-fnvkfyppdpqsgltedtryfclclrqdlasgrpcsf 330
 QY 154 TSAALLSHIVQSEIGDFDEALDRH----LAKNKYIP-QQDALEKIVFEFHHNHIGQTP 208
 Db 331 vthallgsytlqaeigdyd---peehgsidsefqtaptqtkeelkvaehkthrglsp 387

QY 209 AESDFOLLEIARRLEMYGIRLHPAKDRGTGKINIAVANTGILVFGQFTKINAFNNAKVRK 268
 Db 388 aqadsqfllenakrismygdvllhahadsegvdikigvcangllykdrlnrfawpkilk 447
 QY 269 LSEKRELIKLRPDANSAYODTLEFLMASRDFCKSFWKICVEHAFERLFEPEPKPKPK 328
 Db 448 isykrnsfykvrpaelqfestigfkipnhrakraikwkvcehntfyrly-speqppka 506
 QY 329 VLSRGSSFRSGRTQKQVLDYVREGGHHKVKQFERKHSKIHSIRSLASQPTLENSVLEQ 388
 Db 507 kfltlgskfrysgtrtqatqastldrpaphfertskrvs-rslgapi---gymdq 561
 QY 389 SQQSTSLTFGBGASPGQSGCRKEPKVSGAGEGSHPSAPRSPAGNKOADGAASAP- 447
 Db 562 slmk-----dfpagaag-----eisygpgl-vslavvgdgdgrrevrptkaph 604
 QY 448 -----TEEEVVVKDRT-----QOSKPOPP- 467
 Db 605 lqliegknslrvegdnlvyrhslmleeldkagedllkhqasiselkrnfemestpepr 664
 QY 468 -----QPSTGSLTGSPPH-----LSELSVNSQGVAPANVT-----LSPNLSP-DTK 507
 Db 665 newekrritpislqtgssshetlniveekkraevgkderviteemngkeispgsggeir 724
 QY 508 QASPLI-----SPLNDQACPTDDEGRRKRFTDKAYFIAKEVSTERTYTKDLEVIT 563
 Db 725 kvepvtqkdstslsassesseeedvgeyrphhr---vtegtireeqeeyeeveep 781
 QY 564 SWFQSTVSKEDAMPEA 579
 Db 782 rpaakvvveeeavpea 797

RESULT 5

AAB95258
 ID AAB95258 standard; Protein: 687 AA.
 XX AAB95258;
 AC AAB95258;
 XX 26-JUN-2001 (first entry)
 XX Human protein sequence SEQ ID NO:17435.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX Claim 8; SEQ ID 17435; 2537pp + CD ROM; English.
 XX

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 687 AA;
 SQ

Query Match 11.5%; Score 626; DB 22; Length 687;
 Best Local Similarity 30.1%; Pred. No. 61e-43;
 Matches 191; Conservative 92; Mismatches 221; Indels 130; Gaps 22;

QY 31 PPPTSGKLVSIKIQMLDDTQEAPEVQAPRQKVLDAVCNHLNLVDEGDFGLFPPDHK 90
 Db 34 paagdsksilctcrvslldgtvslpdkakgqelfqimylhldiesdyfglrfmdsaq 93
 QY 91 ITWVLDLLKPIVKQIRPKHVVVVFVFPFDPHTQLOBELTYLFLQVQKDLAQGRLT 150
 Db 94 vahldgtksikkqkigspcyhlrvkfyssennlreeltryflvlgkqdlsgkld 153
 QY 151 CNDTSALLSHIVQSEIGDFDEALDREH---LAKNKYIP-QQDALEKIVFHHNHG 205
 Db 154 cpfdtavqlaaynlqaelgdydia---ehspelvseffvpiqteemelaifekwkyrg 210
 QY 206 QTAESDFQLEIARRLEMYGIRLHPAKDRGTGKINIAVANTGILVFGQFTKINAFNNAK 265
 Db 211 qtpagaetnylnkakwlemygvdmhvkvkardgndyslgltptgylvfgdtkigllfwfk 270
 QY 266 VRKLSFKRRKF-LIKLRPDANSAYOD-TLEFLMASRDFCKSFWKICVEHAFERLFEPEP- 322
 Db 271 itrldfkknkltlvvvedddgqkehtfvridhpkackhlwkavvehhaffrl-rgpv 329
 QY 323 -KPKPKPVLSRGSSFRSGRTQKQVLDYVREGGHHKVKQFERKHSKIHSIRSL-----AS 376
 Db 330 qkshrsrgfirlsgrfrysgktqvtktkn--arrstsferrpskrysrtrtlmqkacat 387
 QY 377 QPTLENSVLEQSQQSTSLTFGBGASPGQSGCRKEPKVSGAGEGSHPSAPRSPAG 436
 Db 388 kpeel-----svhnvnstqsgsqgawmralspvspsis----- 422
 QY 437 NKQADGAASAPTEEEVEVVKDRTQOSKPOPPQ-----PSTGSLTSGPHSELVNS 487
 Db 423 -----sapvpveie-----nlpyspgtdqdrkcipnldilnspdlletti-- 464
 QY 488 QGVVAPANVTLS-----PNLS-----PDTKQA-----SPLISPL 516
 Db 465 -gdvigaasdmetsqalndvnnvatrlpglgepeveyetlkdteeklkqlmenspllspr 523
 QY 517 LN-----DQACPRTDDEDEGRR-KRFFPTD-----KAYFIAKEVSTERTY 555
 Db 524 snldvninsqeevvkltcklennviespglnvrmvppdfksnllkqaqveavhvktdsll 583
 QY 556 L--KDLEVITSWFQSTVSKEDAMP---EALKSLI 584
 Db 584 lshknanvqdaatnsavlnennvplpksetlm 617

RESULT 6

AAW12522
 ID AAW12522 standard; Protein; 913 AA.
 XX AC AAW12522;
 XX DT 22-MAY-1997 (first entry)
 XX DE Protein tyrosine phosphatase that localises to focal adhesion.
 XX KW Protein tyrosine phosphatase; PTPH1; focal adhesion;
 KW protein tyrosine kinase; malignancy; cancer; gene therapy;
 KW retrovirus; vector.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Region 30..357
 FT /note= "N-terminal region shows homology with the
 FT N-terminal regions of the talin family"
 FT Region 357..663
 FT /note= "central region includes sequences with
 FT features of sites of phosphorylation by
 FT casein kinase II and p34cdc2"
 FT Modified-site 372
 FT /label= Phosphorylation
 FT /note= "putative p34cdc2 phosphorylation site"
 FT Modified-site 381
 FT /label= Phosphorylation
 FT /note= "putative p34cdc2 phosphorylation site"
 FT Modified-site 424..428
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 438..442
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 489..492
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 514..517
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 543..547
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 607..610
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Region 664..931
 FT /note= "C-terminal region shows homology to "
 XX US5595911-A.
 XX PN 90US-0494036.
 XX PD 21-JAN-1997.
 XX PF 14-MAR-1990;
 XX PR 01-MAR-1991;
 PR 14-MAR-1990;
 PR 16-AUG-1993;
 XX 93US-0107420.
 XX PA (COLD-) COLD SPRING HARBOR LAB.
 XX Tonks NK;
 XX

DR WPI: 1997-107583/10.
 XX N-PSDB; AAT58627.
 PT DNA encoding protein tyrosine phosphatase - for gene therapy of
 PT cancer
 XX
 XX Disclosure; Fig 1A-B; 12pp; English.
 XX
 CC A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the
 CC dephosphorylation of proteins in which tyrosyl residues have been
 CC phosphorylated through the action of a protein tyrosine kinase
 CC (PTK). It localises to focal adhesions, a major site of action of
 CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)
 CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian
 CC cells in sufficient quantities to overcome or counteract PTK activity.
 CC Phosphorylation of tyrosine residues at abnormal levels is
 CC prevented or reversed, resulting in the prevention or reversal of
 CC malignancy of cells.
 XX Sequence 913 AA;
 SQ
 Query Match 9.9%; Score 541; DB 18; Length 913;
 Best Local Similarity 24.3%; Pred. No. 1.1e-35;
 Matches 232; Conservative 150; Mismatches 346; Indels 228; Gaps 41;
 QY 27 RGQPPPTSGKLVSIKIQMLDDTQBAFEVPPQAPQKVLDAVCNHLNLVEGDFGLFEP 86
 DB 16 rtselepketrseivcsihfldgvvtfkvtkdgtqvlldvmhnlhgvtekeyfglqhd 75
 QY 87 DHKKITV-WDLLKPIVKQIRRPKHVVVVFVFFPDHFTQLQEEELTRYLFALQVQDLA 145
 DB 76 ddsdvdpwleaskpkrklkggfpctllhrvrfpdpntlqgeqrthlyflqlkmdic 135
 QY 146 QGRLTCDNTSAALLISHIVQSEIGDFDEALDRE-HLAKNKYIPQDALED---KIVEFHH 201
 DB 136 egritcplnsavvlasayavqshfgydnssihbpgytsdshfipdq--edfltkvesl 193
 QY 202 NHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTINAF 261
 DB 194 qhsghlkqseascyiniartldfygvelhsgrdldhldmigasagavaykyictsfy 253
 QY 262 NNAKVKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKFWKICVHHAFRLFEE 321
 DB 254 pwnnilksifkrkffihgrkqaesrehivafnmynrsknlwkscevhtffga-kk 312
 QY 322 PKPKPKPVL--FSRGSSFFSGRTOKQLD-YVKE--GG-----HKKVQPERKHSKI 368
 DB 313 llpqeknvlsqywtmgsr-----ntkksvnnqyckkvigmmvwnpamrrsisvehletk- 366
 QY 369 HSIRSLASQPTELNSELVLEQSQGSTSLTFGEAESPGQSCRRGKPKVSAAGEPGSHSP 428
 DB 367 ----slpsrspi-----tpnws-----prl-----rheir 389
 QY 429 APRRSAGNKQADGAASAPT--EEEEVVKDRTOQSKPQPQSPQSTGSLTSGPHLSLVN 486
 DB 390 kprhs-----sadnlanemtyitetedvfytykgslapg-dsdsevsqnrshqeslsen 443
 QY 487 SQGCVAPANVTLSPNLSPTKQASPLISPLLNDOACPTDDEDEGRKRKRPDTPDKAFIAK 546
 DB 444 n-----padsyl-----tqsssvspssnapgscspgvdqql-----lddfhrvtk 486
 QY 547 EVSTTERTYLKDELVITSWFQSTVSKEDAMPEALKSL-IPNPEPLHKTFTNLKETEQR 605
 DB 487 ggsctedas-----qyycdkndngdsylvirirtpded--gkfgnlkggvdqk 532
 QY 606 LALWEGRSNAQ-----IRDYQRIGDVMKLNKIQMKH-----LAAHLWKHSEALEA 650
 DB 533 mplvvsrinpesadtcipklnegdqvlingrdisehtdhqvmfikareshsrel-a 591
 QY 651 LENGIKSSRRLENFCRDFELOKVCYLPNTFLLRLPLHMLHYKQVLERLCKHHPPSHADF 710

Db 592 lvirravrsfadksedelngl-----fpeaifpmc-----peggd- 628

Qy 711 RDCRAALAEITEMVAQLHGTMIKME-----NFQKLHELKLDLIGIDNLVVPGRGF 760

Db 629 -----tlegsmaklkglesgtvliqfeqlrykk-----pglai 662

Qy 761 IRLGSLKSLGKGLQRMFFLNDVLLY-TSRGLTASNQPKVHG-----QLPLYGMTIEE 814

Db 663 ----tfaklpqldknr-----ykdvlpydttrvllqgnedyinasvynmeipaanlvnky 714

Qy 815 SEDEWGVPH-CL-----TLRGOROSIIVAASSRSEMEKWEVDIOWAIDLAEKSSSPAPEFL 869

Db 715 iatgdpiphtcaqfwvwdqklsily-----mlttitergrtkchqy 758

Qy 870 ASSPPD-----NKSPDEATAADOSEDDLSASRTSLEROAPHRGNTMWHVCW 916

Db 759 -pdppdvmmhbgfhiqcsedctiayvsremlvntqtgehtvth-----lqyvaw 809

RESULT 7

AA25156

ID AAY25156 standard; Protein; 913 AA.

XX AC AAY25156;

XX DT 07-SEP-1999 (first entry)

XX DE Human PTPH1 protein.

XX KW PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;

KW localisation; treatment; overexpression; oncogenic; cell transformation;

KW prevention; phosphotyrosine; disease; malignant.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

FT Region 30..337

FT /note= "region of homology to the N-terminal domain

FT of band 4.1, ezrin and talin. This region is

FT known to be important for localisation to focal

FT adhesions"

FT Modified-site 372

FT /note= "potential p34cdc2 phosphorylation site"

FT Modified-site 381

FT /note= "potential p34cdc2 phosphorylation site"

FT Modified-site 424..428

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 434

FT /note= "potential p34cdc2 phosphorylation site"

FT Modified-site 438..442

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 489..492

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 514..518

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 543..547

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 607..610

FT /note= "potential casein kinase II phosphorylation site"

FT

PN US5863781-A.

XX

XX 26-JAN-1999.

XX 04-DEC-1996; 96US-0759536.

XX

PR 01-MAR-1991; 91US-0663579.

PR 14-MAR-1990; 90US-0494036.

PR 16-AUG-1993; 93US-0107420.

PR 04-DEC-1996; 96US-0759536.

XX

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Tonks NK;

XX WPI; 1999-131308/11.

DR N-PSDB; AAX78463.

XX

PT Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells

XX

PS Claim 1; Fig 1A-B; 12pp; English.

XX

CC This sequence represents a novel protein tyrosine phosphatase, PTPH1,

CC isolated from HeLa cells. The protein of the invention appears to

CC localise to focal adhesions and is therefore potentially useful in the

CC treatment of cancer. Overexpression of PTPH1 can be used to counter the

CC effects of oncogenic protein tyrosine kinases such as those of

CC transforming viruses and for interfering with or reversing cell

CC transformation. This would provide a means of preventing or reversing

CC abnormally high levels of phosphotyrosine associated with any disease or

CC condition such as preventing or reversing malignancy associated with the

CC activity of a protein tyrosine kinase.

XX

SQ Sequence 913 AA;

Query Match 9.9%; Score 541; DB 20; Length 913;

Best Local Similarity 24.3%; Pred. No. 1.1e-35;

Matches 232; Conservative 150; Mismatches 346; Indels 228; Gaps 41;

Qy 27 RGOKPPPTPSGKLVSIKIOMLDDTOEAFEPORAPGKVLDDAVCNHNLNVEGDYFGLFEP 86

Db 16 rtseipektrsevischfhdgvgvqtkqtdgtgvlidmwhnhlgvtekeyfiglhd 75

Qy 87 DHKKITV-WLDLKLPIVKQIRRPKHVVVKEVAFPPDPHTQLQOEELTRYLFALQVKODLA 145

Db 76 ddsavdsprwleaskpirkqlkggfpctlhfrvrfipdpntlqqeqtrhlyfqlkmdic 135

Qy 146 QGRLTCDNTSAALLISHIVOSEIGDDEALDRE-HIAKNYIFQDQDALED---KIVFEHH 201

Db 136 egrltcpnlssavvasvayqshfgydnssihhpgyisdshfipdn--edfltkveslhe 193

Qy 202 NHGQTPAESDFQLLETARLEMGVIRLHPAKDREGTKINLAVANTGILYFQGTFTKINAF 261

Db 194 qhsglkqaseacycinlartldfygvelhsgrdhldmldmglasgavayrvkyictsfy 253

Qy 262 NNAVKRLSKFRKRFLTKLRPDANSAYQDTLEFLMASRDFCKSWKICVHHFAFFRLEFEE 321

Db 254 pwnilkisfrkkfihqkqaesrehivafnmlyrskulwkcscvehhtffga-kk 312

Qy 322 PKPKPKPVL---FSRGSSFRSGRTQKVLD-YVKE---GG-----HKKVQPERKHSKI 368

Db 313 llpqeknvlsgywtmgsr-----ntkksvnnqyckkvigwmvnpamrrralsvehletk- 366

Qy 369 HSIRSLASQPTLNSEVLEQSQOSTSLTFGEAESGGQSCRRGKPKVKSAGEPGSHSP 428

Db 367 ----slpsrppl-----tpnwr-----prl-----rheir 389

Qy 429 APRRSPAGNKQADGAASAPT--EEEEVVKDRTQQSKPQPQPSTGSLTSGPHSEL SVN 486

Db 390 kprhs-----sadlanemyitetedvfytykgslapq-dsdsevsqnrspqhesl sen 443

Qy 487 SQGGVAPANVTLSFNLSPDTKQASPLISPLNDOACPRTDDEGRKRKRPTTKAYIAK 546

Db 444 n-----padsyl-----tqksssvspssnagpcspgdvqdql-----lddfhrvtk 486

Qy 547 EVSTTERTYTKDLEVITSWFQSTVSKEDAMPEALKSI-IPFNPEPLKFTNFKLKEIOR 605

Db 487 ggsteda-----qyycdkndngdsylviriltpded--gkfyfnlkgvgvdqk 532

Qy 606 LALWEGRSNAQ-----IRDYQRIQIDYMLKNIQGMKH-----LAAHLWKHSEALEA 650

Db 533 mpivvsvrinpespadtcipklnegdqivlingdisethdgvvmfikasreshrel-a 591

Qy 651 LENGIKSSRRLENFCRDFELQKVCYLPNTFTLLRPLRLHLMHYKQVLERLCKHPPSHADP 710

Db 592 lvirrravrsfadksedelngl-----fpeaifmc-----peggd- 628
 QY 711 RDCRAALAEITEMVAQLHGTMIKME-----NFQKHLKLDLIGDNLVPGREF 760
 Db 629 -----tlegsmaglkglsgtvlilqfeqlryrk-----pglai 662
 QY 761 IRLGSLKSLGKGLQRMFFLNDVLLY-TSRGLTASNOFKVHG-----OLPLYGWTIEE 814
 Db 663 ----tfaklpqnlknr----ykdvlpydtrvlllgnedyinasvymneipaanlvnky 714
 QY 815 SEDEMGVPH-CL---TLRGQSIIVAASSRSEMEKWEVDIQMAIDLAEKSSPAPEL 869
 Db 715 iatqgpiptcaqfwvwdqkisiiv-----mltltergrtkchgyw 758
 QY 870 ASSPPD-----NKSPDEATAAQESEDLLSASRTSLERQAPHRGNTMVHVCW 916
 Db 759 -pdpdvmmhggfhicqsedctiayvstremlvntqtgeehvtth----lgyvaw 809
 RESULT 8
 AAY51248
 ID AAY51248 standard; Protein; 766 AA.
 AC AAY51248;
 XX
 DT 14-APR-2000 (first entry)
 DE Rat actin-binding protein frabin.
 KW Actin-binding protein; frabin; rat; developmental stage; animal growth.
 XX Rattus sp.
 XX JP11346775-A.
 PD 21-DEC-1999.
 XX
 PF 11-JUN-1998; 98JP-0164016.
 XX
 PR 11-JUN-1998; 98JP-0164016.
 XX
 PA (KAGA-) KAKAKU GIJUTSU SHINKO JIGYODAN.
 PA (OBAL/) OBAISHI H.
 XX
 DR WPI; 2000-109691/10.
 DR N-PSDB; AA244678.
 XX
 PT An actin-binding protein frabin - plays an important role in the
 PT developmental stages of animal growth
 XX
 PS Claim 1; Page 6-8; 12pp; Japanese.
 XX
 CC This invention describes a novel actin-binding protein, frabin (I). The
 CC new actin-binding protein frabin plays an important role in the
 CC developmental stages of animal growth. The protein can be used in gene
 CC engineering. This sequence represents the rat frabin protein described
 CC in the method of the invention.
 XX
 SQ Sequence 766 AA;
 Query Match 9.5%; Score 520.5; DB 21; Length 766;
 Best Local Similarity 23.5%; Pred. No. 4.2e-34;
 Matches 180; Conservative 152; Mismatches 321; Indels 113; Gaps 22;
 QY 364 KHSKIHSIRLASQPTLSENVLEQSQQTSLTFGGEAESPQGQSRGKPKVSGAGEPG 423
 Db 16 khsksdlisfhfeggsydlstvdqkdstmnlipdtrghgitsttpqklp----- 67
 QY 424 SHPSAPRRSPAGNKQ-----ADGAASAPTEEEVVKDRTOQSKPQPPOPTGSLTGS 477
 Db 68 shkspkqkedsdngqhgclangvaqaqsmce-----tekealspetdtqtaas 122

QY 478 P--HL-----SELVSNQSGGVAPAN-----VTLSPNLSPTDKOASPLISPLNDQA 521
 Db 123 pdahvngvrntttdssassvtnshdenacdsctrggtdlglpskegepvieaelqere 182
 QY 522 CPTDD-----EDEGRKRFPDTKAYFIAKEVSTTERTYVKLDLEVITSFQSVSRKE--- 573
 Db 183 nglsteglapldqhhevketneqklhkiatelllterayvvrlnlildqvfykilleanr 242
 QY 574 DAMPEALKSLIFPNFPLHKKFTNF-LKETEORLALWEGRSNAQIRDYORIGDVMKLNQ 632
 Db 243 gsfpaemvknifsnissinafshkfillpelekrmgewe-----ttrigdlqklap 294
 QY 633 GNMKHLAAHLWKHSEALEALENGIKSSRRLENFCRDELQKVC-YLPLNPLFLPLRLHLMH 691
 Db 295 flkmygeyvgfdnavelvknmtervpqtksvteeiqkqkicgsltlqhmlepiqrpr 354
 QY 692 YKQVLERLCKHHPPSHADFRDRAALAEITEMVAQLHGTMIKMFQKHLKLDLIGD 751
 Db 355 yemilkdyllkispdpdwndakksleistaashnsairkmenkkllleiemigeee 414
 QY 752 NLVVPGRFIRGLSLSKLSGK--GLQORMFFLNDVLLY-TSRGLTASNOFKVHGQLPLY 808
 Db 415 divnpsnelikeqilklaarntsaderylflfnmllcyvprfslvgsktvtrvgid 474
 QY 809 GMTIESEDEWGVPHCLTLRGQRQSIIVAASSRSEMEKWEVDIQMAIDLAEKSSPAPEF 868
 Db 475 gmkivethne-eyphtfvsggkterlclgasseqdkeewikaigesidafrhetfrna 533
 QY 869 LASSPPDNKSPDEATAAD-----QSEDDL-----SASRTSLERQAPHRGNTMVHVC 915
 Db 534 iak---endiplestvtaelgraprwrindvntcmkckesfnaltrrrhcracghvc 590
 QY 916 WHRNTSVSMVDFS-----IYVEN-QLSGN-----LL 940
 Db 591 wkcsdykaqlaydgrlnkvckdcyqimgfaeseekrrgileiesavgsevcsl 650
 QY 941 RKFKNNGWQKLWVFTN---FCLFFYKSHQDNHPLASLLGLYSLTIPSESENIQKDYV 997
 Db 651 qymekskpwqkiwcviipkqdpvlvmygapdvraqatipllgyivddmpksadl--phs 708
 QY 998 FKLHFKSHVYFRAESEYTFERWMEVIRSATSSASRPHVLSHRESL 1043
 Db 709 fkitqsksvhsfaadseelkqkwikilllavtgetpgpsehldtl 754
 RESULT 9
 AAU04486
 ID AAU04486 standard; Protein; 369 AA.
 XX
 AC AAU04486;
 XX
 DT 26-SEP-2001 (first entry)
 DE Human protein tyrosine phosphatase (PTPase) polypeptide #2.
 XX
 KW Protein tyrosine phosphatase; PTPase; human; nervous system; dementia;
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; autism;
 KW Tourette's syndrome; obsessive compulsive disorder; schizophrenia; mania;
 KW cardiovascular disorder; atherosclerosis; myocarditis; Addison's disease;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; vitiligo;
 KW ankylosing spondylitis; Sjogren's syndrome; asthma; atopic dermatitis;
 KW diabetes mellitus; inflammatory disorder; pancreatitis; sarcoidosis;
 KW allogeneic transplant rejection; multiple myeloma; Burkitt's lymphoma;
 XX Leukaemia; cancer; neoplasm.
 OS Homo sapiens.
 XX
 PN WO200153530-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01563.


```

Db 126 sgmyklnmlslagmkvkrptqe-ayqnelkiesversflissassaterdewieaisrai 184
QY 856 -DLARKSSPAPEFLASSPPDNKSPDEATAADQSEDDLSASRTSLERQAPHRGNTMVHV 914
Db 185 eeyakkritfcp-----srside---adsenkeevspgskapiwipdtratcmi 232
QY 915 C-----WHR----- 918
Db 233 ctseftltwrthracgkivcacsnnkygldylnqparvcehcfcgqlqldhghspr 292
QY 919 -----NTFSVMVDFSI-----AVENOLSGNLLRKFKNNGW 949
Db 293 lqspgnhkspsalsvshsipsgrkqkkipaalkevsantedssmsgylyrskgnkxpw 352
QY 950 QKLWVFTNFCULFFKSHQDNHLASPLILGYSLYTPSESENIQDYVFKLHFKSHVYV- 1008
Db 353 khfwfvknkvlyaaesqpllgftv-iqvnden-seskvfqlhknmlfyv 410
QY 1009 FRAESEYTFERWMEVIRSAT 1028
Db 411 fkaedahsaqkieafqegt 430

RESULT 11
AAW67438
ID AAW67438 standard; Protein; 1174 AA.
AC AAW67438;
DT 02-MAR-1999 (first entry)
XX Human protein tyrosine phosphatase D1 protein.
DE Human; protein tyrosine phosphatase; skeletal muscle; hybridisation;
KW glycoprotein; probe.
XX Homo sapiens.
OS Homo sapiens.
FH Key
FT Domain
FT Binding-site
FT Binding-site
FT Binding-site
FT Binding-site
FT Modified-site
FT Modified-site
FT Modified-site
FT Modified-site
FT Modified-site
FT Binding-site
FT Modified-site
FT Region
FT Domain
FT
XX US5831009-A.
XX
XX 03-NOV-1998.
XX
XX 22-MAY-1995; 95US-0446345.
XX

```

```

PR 27-APR-1994; 94US-0234440.
PR 15-AUG-1992; 92US-0923740.
XX 22-MAY-1995; 95US-0446345.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX Moller KB, Moller NPH, Ullrich A;
XX WPI; 1998-609316/51.
DR N-PSDB; AAV34368.
XX Protein tyrosine phosphatase D1 - useful in screening assays for
XX e.g. agonists or antagonists
XX Claim 1; Fig 5A-J; 61pp; English.
XX This is the sequence of the complete human protein tyrosine phosphatase
XX D1 (PTP-D1). The coding sequence was isolated from a human skeletal
XX muscle cDNA library. The phosphatase is used e.g. in a hybridisation
XX assay for detecting a nucleic acid encoding a normal or mutant PTP-D
XX protein or glycoprotein in a cell or subject, using a nucleic acid
XX comprising at least part of a normal or mutant PTP-D protein or
XX glycoprotein as a probe; or a screening assay for a compound capable
XX of binding to a PTP-D protein or glycoprotein, using a PTP-D protein
XX or glycoprotein or the compound-binding portion of it attached to a
XX solid support.
XX Sequence 1174 AA;
SQ
Query Match 7.2%; Score 396; DB 19; Length 1174;
Best Local Similarity 22.5%; Pred. No. 1.8e-23;
Matches 141; Conservative 103; Mismatches 248; Indels 134; Gaps 18;
QY 25 LERGOKPPPTSGKLVSIQIMLDDTQEAPEVPQAPGKVLDDAVCNHNLNVEGDYFGL 84
Db 9 lkrtryvssksclva-riqlnnfevftisvestgqesleavaqrlelrevtyfslw 67
QY 85 FPDHKKITVMDLLKPIVKQI-RRPKHVYVVFVFPDPDHTQLQELTRYLFAQVKOD 143
Db 68 yynqknqrwvdklepkldkyaaleptvfygvvfpvsvsqqleltryqyylqkkd 127
QY 144 LAQRLTCNDTSAALLSHIVQSEIGDFDEALDREHLAKNKYIP-----QDQADEL 195
Db 128 ilsgisptcleqqlaglavdaedfgdqesqdfqlkfalfpvgwldkveleatqk 187
QY 196 IVEFHNNHIGQTPAESDFOLLETARLEMYGIRLHPKADREGTKINLAVANTGILV-FOG 254
Db 188 vallhqkyrgltapdaemlymgevermdygeesypakdsqgsdisigacleqifkxkn 247
QY 255 FTKINAFNMAVKRLSKFRKRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHA 314
Db 248 grhpvvrwhdianmshksffalel---ank--eetiqlqtedmetakyvrlcvarhk 302
QY 315 FPRLEF-----EPKPKPKPVLFSGSSFRFSGR-----TQKQVL 348
Db 303 fyrlnqlnqltqtvtvnpirrrssrmslpkbpqyvmpppqllhngyhtepvassqdn 362
QY 349 DVVKEGG---HKVKQPER-----KSKTHSIRS-----LASOPTEL 381
Db 363 fvpnqngyichsqtldraqidfngringvysahstnlnnpqylpapsmnspsit 422
QY 382 NSEVL-----EQSQQSTSLTFG----- 398
Db 423 gsdvmpdylpshrsavippsyrtpdyetvmkqlnrglvhaerqshsrlnigssya 482
QY 399 -----BGAESPGQSCRRGKEPKVSGAGEGSHPSAPRPSAGNKAQDA 443
Db 483 ysrpaalvysqpeirhaqlpsaaahcfsisysfshspypypaerppvvgavvpe 542
QY 444 ASAPTEEEVEVKD--RTOGSKPQPPQSTGSLTGSPHLS-ELSVNSQGVGAPANVTLS 500
Db 543 tnaqlaqagpdpnrmrtqvyrvpppprpanstpdlsrhllyssn---pdltirrv 599

```

QY 501 NLSPTTKQASPL-ISPLNDQACPT 525
 Db 600 hhsvqtfqedslpvahslqevsepl 625

RESULT 12
 AAY34158
 ID AAY34158 standard; Protein; 1174 AA.
 AC AAY34158;
 XX 09-NOV-1999 (first entry)
 DT Human protein tyrosine phosphatase, PTP-D1, protein sequence.
 DE PTP-D1; human; protein tyrosine phosphatase; neoplastic formation;
 XX KW PTPase; growth factor signal transduction; cell cycle progression;
 KW cancer; diabetes; cellular phosphotyrosine metabolism.
 XX OS Homo sapiens.
 XX PN US955592-A.
 XX PD 21-SEP-1999.
 XX PF 27-APR-1994; 94US-0234440.
 XX PR 27-APR-1994; 94US-0234440.
 XX PS 03-AUG-1992; 92US-0923740.
 XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX PI Moller KB, Moller NPH, Ullrich A;
 XX DR WPI; 1999-539599/45.
 XX DR N-PSDB; AA211221.
 XX PT Nucleic acids encoding protein tyrosine phosphatase PTP-D1
 XX useful for identification of PTP-D1 modulators
 XX Claim 1; Fig 5; 63pp; English.
 XX This sequence is the protein tyrosine phosphatase of the invention,
 CC designated PTP-D1. The new PTP proteins are a subfamily of protein
 CC tyrosine phosphatases (PTPases) involved with growth factor signal
 CC transduction, cell cycle progression, and neoplastic formation. The DNA
 CC sequence may be used in the recombinant production of PTP-D1 according to
 CC standard DNA methodology. The protein expressed may then be used in
 CC assays to identify modulators of its enzymatic activity and in the
 CC production of antibodies. The nucleic acids may also be used in assays to
 CC detect and quantify expression of PTP-D1 in samples. Methods identifying
 CC normal or mutant PTP-D genes, or for measuring amount or activity of the
 CC protein can serve as methods for identifying susceptibility to cancer,
 CC diabetes, or other disorders associated with alterations in cellular
 CC phosphotyrosine metabolism.
 XX SQ Sequence 1174 AA;

Query Match 7.2%; Score 396; DB 20; Length 1174;
 Best Local Similarity 22.5%; Pred. No. 1.8e-23;
 Matches 141; Conservative 103; Mismatches 248; Indels 134; Gaps 18;

QY 25 LERGOPTTPSKLYSIKIMLDLDTQAEFEVQAPAGKVLDAVCNHLNVEGDYFGL 84
 Db 9 lkrtrrvtsksclva-rlqlinnfevtlsvestgqesleavagrlreftyislw 67
 QY 85 FPDHKITVWLDLKPVKOI-RRPKHVVKVFFVPPDPHTQOELTRYLFALQVKOD 143
 Db 68 ynnkqqrwldleplkqldkyleptvyfgyvsvsqleqitryqylqkld 127
 QY 144 LAOGRUTCNTSAALLISHIVQSEIGDFEALDREHLAKNKYIP-----QQDALED---K 195

Db 128 ilegspctleqaiqlaglavqadfgdyesqdfqlqkfallpvgwldqekvleatqk 187
 QY 196 IVEFHNNHIGQTPAESDFQLLEIARLEMYGIRLPAKREGTKINLAVANTGILV-FQG 254
 Db 188 vallhkyrgltapdaemlymqevermdgygeesypakdsqgsdisigaclegifvkhn 247
 QY 255 FTKINAFNAKVRKLSFKRRKFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHA 314
 Db 248 grhpvvrwhdianmhsksfalel---ank--eeicqfqtmedetakiwricvarhk 302
 QY 315 FFLRFE-----EPKPKPKPVLFSGSSFRFSGR-----TOKQVL 348
 Db 303 fyrlncnqlqtvtvnprrrrssrmslpkqpymppppqlhynghtyepvassqdn 362
 QY 349 DYVKEGG---HKVQFER-----KHSKIHSIRS-----LASOPTEL 381
 Db 363 fvpngngyysqtsldraqidfngrirngsvysahstnlnnpqylqpsmssnpsit 422
 QY 382 NSEVL-----EQQSQSTSLTFG----- 398
 Db 423 gsdvmpdylpshrsavippsyrptdyetvmkqlnrglvhaergshsrlninissya 482
 QY 399 -----EQAESPQGGQSCRCGKEPKVSAEGPSHPSPAPRRSPAGNKQADGA 443
 Db 483 ysrrpaalvysqpeirehaqlpsaaahcftslsyshspypypaeripvgavsvpel 542
 QY 444 ASAPTEEEVEVKD--RTQOSKPPQPPQSTGSLTSGPHLS-ELSVNSQGVAPANTLSP 500
 Db 543 tnaqlgaqdypspnlmrtqvyrrpppypprpanstpdlsrhllyissn---pdlltrrv 599
 QY 501 NLSPTDKQASPL-ISPLNDQACPT 525
 Db 600 hhsvqtfqedslpvahslqevsepl 625

RESULT 13
 AAB53356
 ID AAB53356 standard; Protein; 635 AA.
 XX AAB53356;
 AC AAB53356;
 XX 09-MAR-2001 (first entry)
 DT Human colon cancer antigen protein sequence SEQ ID NO:896.
 DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.
 XX OS Homo sapiens.
 XX PN WO200005351-A1.
 XX PD 21-SEP-2000.
 XX PF 08-MAR-2000; 2000WO-US05683.
 XX PR 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI WPI; 2000-587534/55.
 DR N-PSDB; AAC98113.
 XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon

PT disorders such as colon cancer -

XX Claim 11; Page 1449-1451; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antinefactive and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX Sequence 635 AA;

Query Match 6.7%; Score 367.5; DB 21; Length 635;
Best Local Similarity 24.0%; Pred. No. 1.6e-21;
Matches 125; Conservative 98; Mismatches 226; Indels 71; Gaps 16;

QY 7 RPPGSRGAPEN-----SGISTLERGOKPPPTSGKLVSIIQIQLDDTQEAFFVQR 59
DB 19 rstagaglgqrccwvifsgldchqpkmp-----kpinvrvtmdaele-faigpn 72
QY 60 APGKVLDAVCHNLNVEGDYFGLFEPDHPKKTITVWDLKPI-VKQIRPKHVVVVVK 118
DB 73 ttgkldqdvktigirevwyfghydnkgfptwklkdkksaqevrkenplqfkfrak 132
QY 119 FPPPD-HTQLOELTRYLALQVKQDLAQRGRLTCNDTSAALLSHIVQSEIGDFDALDR 177
DB 133 fypedvaeellditqklfflgvkegildseiyccpbtavilgsvayqakgdynevhk 192
QY 178 E-HLAKNKYIPQ-----QDALEDKIVERHHNHIGOTPAESDFQLLEIARRLEMYGI 227
DB 193 sgyislerlipqrvmdghkltrdqwedrigvwhaehrgmlkdnamleykiaqglemygi 252
QY 228 RLHPAKDREGTKINLAVANTGILVFGFTKIN---AFNNAKVRKLSFKRRKRLI----KL 280
DB 253 nyfeiknkgtdlwgvdalgnlyekddkltkpgifpseirnisfndkfkvikpidkk 312
QY 281 RPDANSAYQDTLEFLMASRDFCSFWKICVEHHAFFRLPEEPKPKPKPVLFSGSSFRFS 340
DB 313 apd-----fvfayprlrinkrilqlcmgnhellymrrrkpdrtievqgmkaqareekhq 364
QY 341 GRTQKQVLDVYKE-----GGHKKVQFERKHSKTHSIRLSASOPTELNSEVLEQSQSTSL 395
DB 365 kqiergqletekkrretverekeqmmrekeelmrlriqdyeeekckkaerelseiqralql 424
QY 396 -----TFEGAE--SPGGQSCRRGKE-----PKVSAGEPGSHPSAPRRSPAG 436
DB 425 eeerakraqeaeerleadmaarakeelerqavdqiksqeqlaaleaeytakialleear 484
QY 437 NKQADGAAS---APTDEEEVVKDRTO-----QSKPPQPPQ 469
DB 485 rrkedeveeqhakeaqdvlvkteelhlvmtpapppppp 524

RESULT 14

AY27443

ID AAY27443 standard; protein; 586 AA.

XX AC

XX AAY27443;

XX XX

DT 26-NOV-1999 (first entry)

XX

DE Amino acid sequence of human ezrin polypeptide.
XX Pharmaceutical; ezrin; mutant; tumor; metastasis; human.
XX Homo sapiens.

OS Key Location/Qualifiers

FX Misc-difference 354
FT /note= "the Tyr at this position can be mutated
FT (preferably to a Phe) to construct an
FT ezrin mutant of the invention"

PN WO9947150-A2.

XX 23-SEP-1999.

XX 18-MAR-1999; 99WO-EP02054.

XX 18-MAR-1998; 98US-0040725.

XX (CURI-) INST CURIE.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Arpin M, Crepaldi T, Gautreau A, Louvard D;

XX WPI; 1999-561851/47.

XX New composition for prevention and treatment of tumors and metastasis
PT
XX
XX Example 1; Fig 1; 31pp; English.
XX
CC The invention provides a pharmaceutical composition containing ezrin
CC protein, RNA or DNA mutated on tyrosine 353, or a functional fragment
CC or derivative of the ezrin mutant. The new composition is useful for
CC prevention and/or treatment of tumors, and especially metastasis. The
CC present sequence represents the amino acid sequence of human ezrin
CC (before the maturation by deletion of the first amino acid Met).
XX
SQ Sequence 586 AA;

Query Match 6.5%; Score 356.5; DB 20; Length 586;

Best Local Similarity 23.9%; Pred. No. 1.1e-20;

Matches 115; Conservative 96; Mismatches 212; Indels 59; Gaps 14;

QY 38 KLVSIKIQLDDTQEAFFVQRPAPGKVLDAVCNHLNVEGDYFGLFEPDHPKKTITVWDL 97
DB 3 kpinvrvtmdaele-faigpnttgkldqdvktigirevwyfghydnkgfptwkl 61
QY 98 LKPI-VKQIRPKHVVVVVKVVFPPD-HTQLOELTRYLALQVKQDLAQRGRLTCNDTS 155
DB 62 dkksaqevrkenplqfkfrakfypedvaeellditqklfflgvkegildseilcyppt 121
QY 156 AALLSHIVQSEIGDFDALDR-HLAKNKYIPQ-----QDALEDKIVERHHNHIG 205
DB 122 avilgsvayqakgdynevhksgyislerlipqrvmdghkltrdqwedrigvwhaehrg 181
QY 206 QTPAESDFQLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTKIN---AFN 262
DB 182 mlkdnamleykiaqglemyginyfeiknkgtdlwgvdalgnlyekddkltkpgifp 241
QY 263 WAKVRKLSFKRRKRLI----KLRPDANSAYQDTLEFLMASRDFCSFWKICVEHHAFFRL 318
DB 242 wseirnisfndkfkvikpidkkap-----fvfayprlrinkrilqlcmgnhellym 293
QY 319 FEEPKPKPKPVLFSGSSFRFSRGTOKQVLDYKE-----GGHKKVQFERKHSKHSIRS 373
DB 294 rrpdpdtievqgmkaqareekhqkqiergqletekkrretverekeqmmrekeelmrl 353
QY 374 LASOPTELNSEVLEQSQSTSL-----TFEGAE--SPGGQSCRRGKE----- 414
DB 354 yeeektkaeerelseiqralqleerakraqeaeerleadmaarakeelerqavdqiks 413

QY 415 PKVAGEPGSHPSAPRRSPAGNKQADGAAS---APTEEEVVVKRTQ-----QSKPQP 467
 Db 414 qeqlaaelaeytakialleearrrkedeveeqhrrakeadldlvktkeelhlymtapppp 473
 QY 468 QP 469
 Db 474 pp 475

RESULT 15
 AAM25892
 ID AAM25892 standard; Protein; 305 AA.
 XX AC AAM25892;
 XX DT 16-OCT-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:1407.
 XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.
 XX OS Homo sapiens.
 XX WO200153455-A2.
 XX PN 26-JUL-2001.
 XX PD 22-DEC-2000; 2000WO-US35017.
 XX PF 23-DEC-1999; 99US-0471275.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX WPI: 2001-457603/49.
 XX N-PSDB; AAH99833.
 XX PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX Claim 20; Page 287; 1217pp; English.
 XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,

CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX Sequence 305 AA;
 SQ

Query Match 6.3%; Score 342.5; DB 22; Length 305;
 Best Local Similarity 33.1%; Pred. No. 5.9e-20;
 Matches 98; Conservative 48; Mismatches 109; Indels 41; Gaps 13;

QY 166 SEIGDFDEALDREH----LAKNKYIP-QQDALEDKIVEFHNNHIGQTPAESDFQLLEIAR 220
 Db 8 aelgdydla---ehspelysefvpigteemelaifekwkeyrggtapaetnynink 64
 QY 221 RLEMYGIRLHPAKDREGTKINLAVANTGIIIVFOGFTKINAFNKAVKRLSFKKRF-LIK 279
 Db 65 wlemygvmhmvkardgndysigltptgvlfegdkiglfwpktrldfknkkitlv 124
 QY 280 LRPDANSAYQD-TLEFLMASRDFCKSFWKTCVEHHFAFFRLFEFP--KPKPKPVLFSGSS 336
 Db 125 vedddgqkeqhtfvrlidhpkackhlwkavehhaffrl-rgpvqkshrsgrfirlgr 183
 QY 337 FRFSGRTQQLVDYVVEGGHKKVQFERKHSKIHISRL-----ASQTEUNSEVLEQSQ 391
 Db 184 frysgkteygttktnk--arrstsfrrpskysrrtlgmakatkeel-----svh 234
 QY 392 STSLTFGEAESPGGOSCRRGKPKVSAGEPGSHSPAP-----RRSPAGNKQAD 441
 Db 235 nnvstqsgsqgawmralspvspsiss-----apvpveienlpqsp-gtdqhd 282

Search completed: December 6, 2001, 08:51:40
 Job time: 332 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 08:37:48 ; Search time 4659.49 Seconds
(without alignments)
12186.589 Million cell updates/sec

Title: US-09-555-342A-1
Perfect score: 3442
Sequence: 1 ccccgagccgcgcgcgtg.....agtattaaacattgtcatt 3442

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3442	100.0	3442	9	AB008430 Homo sapi
2	1054	30.6	2632	10	BC004009 Mus muscu
3	1037.4	30.1	3997	9	AB018336 Homo sapi
4	1019.2	29.6	3719	10	BC009153 Mus muscu
5	342.8	10.0	2888	9	AL161984 Homo sapi
6	339.2	9.9	181800	2	AL137249 Homo sapi
7	332.2	9.4	483	11	G22662 Human STS W
8	286.6	8.3	1438	9	AL122052 Homo sapi
9	254	7.4	96183	9	AL161896 Human DNA
10	254	7.4	152053	2	AC022669 Homo sapi
11	199.2	5.8	170916	9	AL136300 Human DNA
12	194	5.6	2595	9	AF156225 Homo sapi
13	193.4	5.6	152053	2	AC022669 Homo sapi
14	192.4	5.6	2867	9	HMELIA Human struc
15	192.4	5.6	3043	9	HUMEMP41 Human
16	189.4	5.5	5942	10	MUS41SP Mus musculu
17	187	5.4	216	11	G23578 human STS W
18	187	5.4	2882	10	AF106702 Mus muscu
19	185.2	5.4	6102	10	AB019256 Rattus no
20	183.6	5.3	5855	10	AB019257 Rattus no
21	181.8	5.3	3619	9	BC010674 Homo sapi
22	181.8	5.3	3643	9	M68941 Human prote
23	181.2	5.3	2590	4	AF222767 Bos tauru
24	181.2	5.3	3661	10	AB032828 Rattus no
25	181.2	5.3	4543	10	AB032827 Rattus no
26	180.4	5.2	6263	9	AB002336 Human mRN
27	179.6	5.2	5921	3	DROP41A Drosophila
28	176.6	5.1	2964	10	AF044312 Mus muscu
29	174	5.1	3320	10	AF061283 Mus muscu
30	172.4	5.0	2758	5	XELCSK X. laevis cy
31	169.6	4.9	4336	9	AF027299 Homo sapi
32	166.8	4.8	3451	10	AB032366 Mus muscu
33	165	4.8	4051	10	AF152247 Mus muscu
34	162.4	4.7	393	11	G60059 SHGC-130885
35	161.8	4.7	2717	10	AF177146 Mus muscu
36	161.6	4.7	3309	9	AF069072 Homo sapi
37	161.6	4.7	3387	9	BC006141 Homo sapi
38	161.6	4.7	4446	9	AB023204 Homo sapi
39	157.6	4.6	1973	9	AB070172 Macaca fa
40	154.2	4.5	2677	9	HUMELI Human eryth
41	154	4.5	42014	2	AC018327 Drosophil
42	154	4.5	160710	3	AC009537 Drosophil
43	154	4.5	164713	3	AC007589 Drosophil
44	154	4.5	303209	3	AE003604 Drosophil
45	144	4.2	3984	9	M64572 Human prote

ALIGNMENTS

RESULT 1

AB008430	AB008430	3442 bp	mrna	PRI	13-FEB-1999
LOCUS	Homo sapiens	3442 bp	complete cds.		
DEFINITION	AB008430				
ACCESSION	AB008430.1	GI:2766164			
VERSION	AB008430.1				
KEYWORDS	CDEP.				
SOURCE	Homo sapiens	embryo cartilage chondrocyte	cdna to mrna.		
ORGANISM	Homo sapiens				
REFERENCE	1	(bases 1 to 3442)			
AUTHORS	Koyano, Y., Kawamoto, T. and Kato, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-Oct-1997) to the DBJ/EMBL/GenBank databases. Takeshi Kawamoto, Hiroshima University School of Dentistry, Department of Biochemistry, 1-2-3 Kasumi Minami-ku, Hiroshima, Hiroshima 734, Japan (E-mail: tkawamo@ipc.hiroshima-u.ac.jp, Tel:082-257-5688, Fax:082-257-5629)				
REFERENCE	2	(sites)			

QY 1441 cagcccccgagcccaagcaagcaggtccctgactgagcagtcctcaacttccagatgtgt 1500
Db 1441 CAGCCCCGCGAGCCCAAGCACAGGCTCCCTGACTGCGAGTCTCTACCTTCCGAGTGCT 1500
QY 1501 gtgaactcgaggggagtgagccctcccaactgagccttctcccaactgagcccc 1560
Db 1501 GTGAAGTGCAGGGGGAGTGGCCCTCCCAACGTGACCTTGTCTCCCAACCTGAGCCCC 1560
QY 1561 gacaccaagcagggcctccctgtatcagccgctgctgaaagacagcctgccccgg 1620
Db 1561 GACACCAAGCAGGCTCTCCCTTGATCAGCCGCTGCTGAATGACACAGGCTGCCCCGG 1620
QY 1621 acggacgatgaggtgagggccggaggagagattcccaactgataaagcgtactcata 1680
Db 1621 ACGGACGATGAGGATGAGGGCCGGAGGAGAGATGCCCAACTGATAAAGCGTACTTCATA 1680
QY 1681 gctaaaggaagtgtaccacacgagcgaacatatctgaagatctgaagttatcacttcg 1740
Db 1681 GCTAAGGAAGTGTCTACCAACCGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTCG 1740
QY 1741 tggtttcagagcacagttagcaaaagagcgcctacccggaagcactgaaagttctcata 1800
Db 1741 TGGTTTCAGAGCACAGTGAACAGAGAGCGCCATGCCGGAAGCACTGAAAAGTCTCATA 1800
QY 1801 ttccogaattttgaacctttgcacaaatttcatactaattttctcaaggaattgagcaa 1860
Db 1801 TTCCCGAAATTTGAACCTTTGCACAAATTTTATACATAATTTTCTCAAGGAATTTGAGCAA 1860
QY 1861 cgaactgcccgtggaagcgcctcaatcccaaaatacaagattaccagaagaatcggc 1920
Db 1861 CGACTTGCCCTGTGGGAGGGCCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCG 1920
QY 1921 gatgtcatgtgaagaacattcagggcatgaagcactggcggtcacctgtggaagcac 1980
Db 1921 GATGTCTATGCTGAAGAACATTCAGGSCATGAAGCACTGGCGGCTCACCTGTGGAAGCAC 1980
QY 1981 agcaggccttgaggccctggagaatcgaagatcccggtcccggtgagaaacttc 2040
Db 1981 AGCGAGGCTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGTGGAGAACTTC 2040
QY 2041 tgcagagacttgagtcacagaagtgtgtacctaccgctcaacacttccctcgtgg 2100
Db 2041 TGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCCTCTCGGG 2100
QY 2101 ccactgcaacggctcatgcaactacaagcaggctcctggagcggctgtgcaaacaccacccg 2160
Db 2101 CCACTTGCACCGGCTCATGCACCTACAGCAGTCTCTGGAGCGGCTGTGCAAAACACACCG 2160
QY 2161 ccgagccaacgcgaacttcagggaactccgagccgcttttggcagagatcacgagatgg 2220
Db 2161 CCGAGCCACGCGGACTTCAGGGACTTCGCGAGCGCTTTTGGCAGAGATCAGGAGATGGTG 2220
QY 2221 gcacagctccagctacagatgataagtgagagatttccagaagctgcaggaactcaag 2280
Db 2221 GCACAGCTCCAGGTACGATGATCAAGATGAGAAATTTCCAGAAGCTGCACGAATCAAG 2280
QY 2281 aaagatttgatggcatgacaacttgggtccgggaagggaggttccatccgtcggcg 2340
Db 2281 AAAGATTTGATTTGGCATTTGACAACTTGTGGTTCCGGGAAGGGAGTTATCCCTGGGC 2340
QY 2341 agcctcagaactctcctggggagggtccagcagcgcgatgttcttcgtttcaacgac 2400
Db 2341 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATTTCTTCTCTTCAACGAC 2400
QY 2401 gtccctgtatcacagcggggggtgacgctccactcagtttaaaagtccacgggag 2460
Db 2401 GTCCGTGTATACAGAGCGGGGGTGTACGGCTCCATCAATCAATTTAAAGTCCACGGGAG 2460
QY 2461 ctccgccttatggcatgacagattgagagagcgaagcaggtggggggtgcccaactgc 2520
Db 2461 CTCGCCCTCTATGGCATGACGATTTGAGGAGACCGAAGACAGTGGGGGTGCCCACTGC 2520

QY 2521 ctgacccctccggggccagcggcagtcctcatcatcgtggccgcccagttctcgtgccagatg 2580
Db 2521 CTGACCCCTCGGGGGCCAGCGCAGTCCATCATCGTGGCCGCCAGTCTCGGTCCGAGATG 2580
QY 2581 gagaagtgggttgagacatccagattgaccattggcggagagaagagcagcagcccc 2640
Db 2581 GAGAAGTGGTTGAGGACATCCAGATGGCCATTGACCTGGCGGAGAGAGCAGAGCCCC 2640
QY 2641 gcccttgagtttctctggcagcagcccccctgacaaagtccctctgataaagccaccgg 2700
Db 2641 GCCCTTGAGTTCTCTGGCCAGCAGCCGCCCTGACAAAGTCCCTCTGATGAAGCCACCGCG 2700
QY 2701 gctgacccaggaatcagaagatgacctgagcctcgccacatcgtgagcgcagagcc 2760
Db 2701 GCTGACCAAGGAGTGAAGGATGACCTGAGCGCTCGGCACATCGCTGAGCGCCAGGCG 2760
QY 2761 ccgcaccgcggcaacacaaatggtgcacgtgtgctggcaccgcaacaccagcgtctccatg 2820
Db 2761 CGGCACCGGGCAACACAAATGGTGCACGTGCTGGCACCAGCAACACAGCGTCTCCATG 2820
QY 2821 gtggacttcagcatcgagtgaggaatcagttgtctggaacctgctgaggaataatcaaa 2880
Db 2821 GTGGACTTCAGCATCGCAGTGGAGAATCAGTTGTCTGGAACCTGCTGAGGAATTCAAA 2880
QY 2881 aacagcaacgggtggcagaagctgtgggtgtgtgttcacaaacttctgctgttctctac 2940
Db 2881 AACAGCAACGGGTGGCAGAAGCTGTGGGTGTGTTCACAAACTTCTGCGCTGTCTTAC 2940
QY 2941 aaatcacaccaggaacaaatcctccttgcagccttgcctcgtcgtcgtactcgtcaccc 3000
Db 2941 AAATCACACCAGGACAAATCATCCCTTGGCAGCTGCTCTGCTGGCTACTCTGCTCACC 3000
QY 3001 atccctctgagtcggagaacatccagaagaactacgtgttcaagctgcactcaagtcc 3060
Db 3001 ATCCCTCTGAGTCCGAGAACATCCAGAAAGACTACGTGTTCGAAGCTGCACCTCAAGTCC 3060
QY 3061 cagctctactactcaggcgggaagcagatcacagttcgaaggtggaatggaagtgc 3120
Db 3061 CAGCTCTACTTCTCAGGGCGGAAAGCGAGTACACGTTCGAAAGGTGGATGGAAAGTGC 3120
QY 3121 ccagtgccaccagctctcctcgcgaccccgctgttagcacaagaagctctctgtg 3180
Db 3121 CGCAGTGCACACAGTCTGCTCTGGACCCCGACAGTGTGAGGCCACAAAGAGTCTCTGTG 3180
QY 3181 tattgatggccggacacactcgttttcgcgagtggtgcttctcctggaagacgttctctt 3240
Db 3181 TATTGATGGCCGGACACACTCGCTTTCGCGAGTGGTGTCTTCTTGGAGACGTTTCTCTTT 3240
QY 3241 cttctgtattaatgaagcctggttaaaattacacactgtctgaaaaatcaaaaacatggctt 3300
Db 3241 CTTCTGTATTAAATGAAGCCTGGTAAATTTAACACCTGTCTGAAAAATCAAAAAATGGCTT 3300
QY 3301 ccagcagctctcctgtctccacagcgcgcttttttaaccccgacctctcagcgtttgaa 3360
Db 3301 CCCAGAGCTCTCTCTGCTCTCACAGCCGGCTTTTAAACCCCGACCTCTCAGCGTTTGA 3360
QY 3361 tgaacagcgtctccacactcagtcctcgtggcatccgctggggggtgttctttagctagt 3420
Db 3361 TGAACAGCGCTCCCACTCCAGTCTCTGTCCTGTCGTCCTGGGGGCGCTTCTTCTTAGCTAGTG 3420
QY 3421 ccagtattaaacacattgtcatt 3442
Db 3421 CCAGTATTAAAAACATTTGTCATT 3442

RESULT 2

BC004009 2632 bp mRNA ROD 12-JUL-2001
LOCUS Mus musculus, clone IMAGE:3493093, mRNA, partial cds.
DEFINITION BC004009
ACCESSION BC004009
VERSION BC004009.1 GI:13278387
KEYWORDS house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2632)
Strausberg,R.
Direct Submission
Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 8 Row: p Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source

Location/Qualifiers
1..2632

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:3493093"

/tissue_type="Mammary tumor metastasized to lung. Tumor

arose spontaneously from a senescent normal mammary

(clonal) outgrowth infected with the virus MMTV."

/clone_lib="NCI CGAP_Lu29"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

<1..1263

/codon_start=1

/product="Unknown (protein for IMAGE:3493093)"

/protein_id="AAH04009.1"

/db_xref="GI:13278388"

/translation="VRLKNIQGMKHLAHLMLKHSLEALETSIKGSRLEHFCRDFE

LQKVCYLLENFLRLPLRLHMYKHVLERLCKHHPNADPRCAALAEITEMVAQL

HGTMTKMFNFKLHKKDLIDNLVPGREFIRLGSLSKLSGKLQORFFLFNDV

LLYTSRGITASNQFVHGLPLYGMTIEESEEWGVPCHLIRGQRQSIIVAAASRSE

MEKWMEDIQMAIDLAERKNGPTPELLASPPDNKSPDEATAQDEDDLSASRTSLE

ROAPHRGNTMVHVRSTSVSMVDFSIAVENOLSGNLLRKFKNSNGWQKLWVFTNF

CLFFVKSHQDHPPLASPLGLYSLTIPSESENIHKDYVFKLHFKSHVYFRAESEYTF

ERWMEVIRSATSSASRAHLHSHKESHL"

BASE COUNT 632 a 720 c 671 g 609 t

ORIGIN

Query Match 30.6%; Score 1054; DB 10; Length 2632;
Best Local Similarity 89.0%; Pred. No. 1e-230;
Matches 1138; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 1924 gtcattgtaagaacattcaggagcatgaagcactgagcctgacacgttggaagcacagc 1983
|||||
Db 1 GTCATGCTGAAAACATTTCAGGCGATGAGCACCCTGGCGGCCACCTGTGGAAGCACACT 60
QY 1984 gaggcccttgagccctgaggaatgaatcaagagctcccgcgctggaagaaactctgc 2043
|||||
Db 61 GAGGCCCTGGAGCCCTGGAGACCTCCATCAAGGGCTCGCGGGCTGGAACATTTCTGC 120
QY 2044 agagacttgagctgagagtggtgtacctaccgctcaacacacttctctctggcca 2103
|||||
Db 121 CGAGACTTCGAGTGCAGAAAGGTGTGTACCTGCCCGCTCAACACCTTCTCTCTGCCCGC 180

QY 2104 ctgcacccgctcatgcactacaagcagcaggctcctggagcgtgtgcaaacaccaccacccgcgcg 2163
|||||
Db 181 CTGCACCGGCTCATGCATATATAGCATGTCTCTGGAGAGGCTGTGCAAGCACCACCACCA 240
QY 2164 agccacgcccacttcaggagactgcgagcgcgttttggcagagatcacgagatggtggcca 2223
|||||
Db 241 AACCCAGCCGACTTCAGGAGCTCAGAGTGCCTGCGGAGATCACAGAGATGTTGGCC 300
QY 2224 cagctccacggttacgatgatcaagatggagaatttcagaagctgcacgaactcaagaaa 2283
|||||
Db 301 CAGCTGCACGGGACCATGATCAAGATGGAGACTTCCAGAAGCTGATGATGCTCAAGAAA 360
QY 2284 gattgatggcattgacaactcttggttccgggagggaggttcacgtctgggcagc 2343
|||||
Db 361 GATCTGATCGGATTCACAACTTTGTGATCCAGAGGAGGAGTTTCATCCGCTTGGGCAGC 420
QY 2344 ctacgaactctctcgggggaggggctccagcagcagcagcagcagcagcagcagcagcagc 2403
|||||
Db 421 CTGAGCAAGCTCTCCGGGAAGGGGCTTACGACGCGCATGTTCTCTCTGTTCAACGATGTC 480
QY 2404 ctgctatacacagcagcgggggctgacggcctccaatcaagtttaagttcaacgggcagctc 2463
|||||
Db 481 TTGCTGTATACCGCGGGGACTGACAGCATCTAATCAGTTTAAAGTCCACGACAGCTC 540
QY 2464 ccgctctatggcatgacgattgagagagcgaagcagcagcagcagcagcagcagcagcagc 2523
|||||
Db 541 CCACCTCTATGGCATGACGATCGAGGAGAGTGAGGAGAGTGGGGTGTGCCCCACTGCCCTG 600
QY 2524 accctcggggcgagcggcagtcacatcgttgggcggcagttctcgttcggagagatggag 2583
|||||
Db 601 ACCCTGAGGGGCGAGCGGCAATCTATCATCTGTGGCTGCCAGTTCCTCGGTCAGAGATGGA 660
QY 2584 aagtggttgaggaacacagatgacattgacattgacattgacattgacattgacattgacattg 2643
|||||
Db 661 AATGTGATGAGGACATCCAGATGGCCATGTACCTGGCAGAGAAAGAAATGACCCACC 720
QY 2644 cctgagttcctggcagcagcggcggcggcggcggcggcggcggcggcggcggcggcggcgg 2703
|||||
Db 721 CCGGAGCTCTTGCCAGAGAGCCCTCTGACACAAAGTCCCGGATGACGCCAGCGGCT 780
QY 2704 gaccaggagtcagagagatgacattgagcgcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2763
|||||
Db 781 GACCCAGGATGAGGATGACCTGAGCGCTCTCGGACATCGTGGAGCGCCAGCGCCCT 840
QY 2764 caccgcgcgaacacaaatggtcacgtgtgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2823
|||||
Db 841 CACCGTGGCAACACGATGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 900
QY 2824 gacttcagcatcgagtgagaaatcagttgtcttggaacctgctgaggaattcaaaaac 2883
|||||
Db 901 GACTTCAGCATTTGCTGTGGAGAACAGCTCTCTGGGAACCTTGTGTGAGAAAGTTCAAGAAC 960
QY 2884 agcaacgggtgagcaagctgtgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2943
|||||
Db 961 AGCAATGGCTGGCAGAAAGTGGGTGGTATTACCAACTTGTGCTGTGCTGTGCTGTGCTGTGCT 1020
QY 2944 tcacacaggacaatcatcccttgcacgcctgcctcctgcctgcctgcctgcctgcctgcctgc 3003
|||||
Db 1021 TCACACAGAGACAGTACCCCTTGGCAGCCTGCTCTGCTGGGCTACTCTCCCTCACCATC 1080
QY 3004 cctctgagtcgcgagaacatccagaaagactacgtgttcaagctgcacttcaagtcacacacac 3063
|||||
Db 1081 CCTTCTGAGTCTGAGAACATCCACAAGGACTATGTGTTTCAAGCTGCTATTTCAAGTCCCAC 1140
QY 3064 gctactacttcagggcggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3123
|||||
Db 1141 GTGTACTTACTTCAGGCTGAAAAGCGAATACATATTTGAAAGGTGGATGGAAGTGTATCCGA 1200
QY 3124 agtgccaccagctctgcctcgc 3183
|||||
Db 1201 AGTGCCACAGCTCAGCTCTCGGGCCACATCTTGAGTCAACAAAGAAATCTACCTGTATC 1260
QY 3184 tgatggcgaggacacactc 3201


```

Db      1261 TGACGGTGGACAGCGCC 1278
||||| ||| ||||| ||| |||
RESULT  3
AB018336
LOCUS   AB018336
DEFINITION Homo sapiens mRNA for KIAA0793 protein, complete cds.
ACCESSION AB018336
VERSION   AB018336.1
KEYWORDS  GI:3882306
SOURCE   Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
          SK plus clone:hK05692.
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (sites)
          Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
          Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
          Prediction of the coding sequences of unidentified human genes. XI.
          The complete sequences of 100 new cDNA clones from brain which code
          for large proteins in vitro
          DNA Res 5 (5), 277-286 (1998)
          99087487
          2 (bases 1 to 3997)
          Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
          Direct Submission
          Submitted (08-OCT-1998) to the DDBJ/EMBL/GenBank databases. Osamu
          Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
          Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
          (E-mail:cdmainfo@kazusa.or.jp, Tel:+81-438-52-3913,
          Fax:+81-438-52-3914)
          Location/Qualifiers
              1. .3997
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="hK05692"
              /clone_lib="pBluescriptII SK plus"
              /dev_stage="adult"
              /sex="male"
              /tissue_type="brain"
              118. .3282
              /gene="KIAA0793"
              118. .3282
              /gene="KIAA0793"
              /codon_start=1
              /product="KIAA0793 protein"
              /protein_id="BAA3453.1"
              /db_xref="GI:3882307"
              /translation="MGEIEGTQYVLTQAGMLCAQTQPGVSTLEPGQTLTPRMOEKHL
              HLRKLLDNTMEIFDEPKDGOVLLTOVWKRLNIVECDYEGMEFQNTQSYVWLWLP
              KPIRIORPKNVVRLAVKFFPDGQEQEEYRILFALQKRDLEERLTCADTTA
              ALTSHLQSEIGEDYDTHDREKLVNYPGQOHCLEKLEFHQKHVGQTPAESDFQ
              VLEARIKQSEIGIRFMASDREGTKIQLVASHMGVLVFGQTTKINTFNWSKVRKLISF
              RKRFLLKHPHEGPDYQDTLEFLLSGRDECKFEWKICVEYHTFFRLIDQPKKAKAVF
              FRSQSFRYSGRQKQVDFYKDSGMKRIPEYRRHSKTHTSVRALTDLPKQISIFPE
              GLTPASPSANAFYSLSLPTVPSGLPEFKDSSSLTDPOVSYVKSPAERSGAVA
              GLRDTSAQPLGPPALQPGELSTKSPQSPSRKSPLSFQVPLGPAEGSGSP
              LSPVLSDAGAGMDCSEPRHKRPVADAEYFIVKEILATERYTLKDIEVITWFRSAV
              KEDAMPTQLNSIDYPIEYRHFREVEQRLALWEGPSKAHTKGSORIGLLPI
              RNNRQKFTSYQRHDEVTLEKATKCKLEAVYKEFLQKVCLPLNTFLKPI
              ORLLHYRLLLRCGYSFGPHDYADCHDAKKAITEVTTTLOHLLIRLENLOKLELO
              RDLVGENLAPGREITRGCLHLKTKKQLORMFFLSDMLIYTSKGVAGTSHFRI
              GLPLQGLMVEESDNSWSPHCFTTIAAQKTIIVAASTRLEKEKMDLNSALQAAKS
              GBDTALPDRGTVCTPPRPSNVESLEQDEDDARGVRSLEGHGQHRANTTHWVCV
              RNTSVAGRHSAAVENQSLGKLLRKFKNKSNQWGLWVFTNFCFFYKTHQDDYPLAS
              LPLIGYSVSLPREADGIHKDYVFKQFKSHVYFRAESKYTFERWMEVIQACSSASGR
              APSIVODGQGPSSGLEMGVRGKEE"
          998 a 1132 c 1068 g 799 t

BASE COUNT
ORIGIN

Query Match      30.1%; Score 1037.4; DB 9; Length 3997;

```

Query Match 30.1%: Score 1037.4: DB 9: Length 3997:

Db 1170 AGTAGATTATTTCAAAGACAGAGTGGAATGAAGAGAAATTCATATCAAAAGAGGCACAGCAA 1229
QY 1149 gattcattctatccgagccttgcttcacagcctacagaaactgaattcggagctgtgga 1208
Db 1230 GACCCACAGTCC-----GTTGAGGTCTGACTGACAGACCTACCAA 1271
QY 1209 gcagttctcagcagcagcagccttacatttgagaggtgcccgaatctccagggggccca 1268
Db 1272 ACAGAGCATCTCATCCCGAGGATGAGGACTCTCCCTCCCATCTTCACCGAATGC 1331
QY 1269 gaactgcccggcggaggaagaaacgaagtttccgcggggagccggggtcgaccgccag 1328
Db 1332 CTTTATCTCGCTCTCCCTCCACTCTGGTCCCTCTGGCTCGCCTGAGAGTTTAAAGACAC 1391
QY 1329 ccctgcgcgagagagagcccgccgggttaacagagcggcggagcgcctcggcgc 1388
Db 1392 CAGCAGCTCCTCACAGATCCCGCAGGTTTCCTACGTCAAGTCCAGCTGCAGAGGGCG 1451
QY 1389 caggaggagagagagag--tcgttaaggataggaccagcagagtaaacctcagcc 1445
Db 1452 CAGTGGAGCAGTGGCTGGAGGCCCGACACACCATCGGCCAGCCCTCGGGCCCCCGC 1511
QY 1446 ccgcgagcaagcaagggctccctgactggaagtcctcactcactcactcactcactcact 1505
Db 1512 ACTCCAGCCTGGTCCAGAGCCTTTCCACGAAGATCCTCAGCCTTCTCCCTCCAGCCGGA 1571
QY 1506 ctgcgagggggagtgccctcgaacgtgacctgtctcccaacctgagcccgacac 1565
Db 1572 GAGCCCCCTGAGTCTGAGCCCTGCATTTTCAGGTGCGCTT-----TGGGCCCACTGA 1622
QY 1566 caagcaggcctctcccttgatcagccgcctgtgtaagaccagcgctgccccggagcga 1625
Db 1623 ACAGGGCTCATCCCACTCTCTGAGCCCTGTCTCTAGTGATGTGGCGGAGCCGGGATGA 1682
QY 1626 cgaatgagatgagggccgaggaagagattcccaactgataaagctactcactaactaa 1685
Db 1683 CTGCGAG--GAGCCAGACACAAAGCCGCTGCTGAGAGCAGAGGCCCTACTTCTATGTC 1739
QY 1686 ggaagtgtctacaccagcagcaatctgtaaggtatctcgaagttatcactctggtt 1745
Db 1740 AGAGATTCTCGTACAGACGACATACCTCAAGGATTTAGAGTTATTTACCGTGTGGTT 1799
QY 1746 tcagagcacatgagcaagagagccatgcccggagcactgaaagtctcattatcc 1805
Db 1800 CCGCAGCGAGTGGTGAAGGAGGAGCCCATGCTCGGACTCTGATGACGCTGCTCTCTC 1859
QY 1806 gaatttgaaactttgcacaaatttctaactaattttctcaaggaattgagcaacact 1865
Db 1860 CAACATCGATCCCATCTATGAGTTCACAGAGCTTCTTGGCGAGGTTGGAGCAGAGGT 1919
QY 1866 tgccttggtggaaggccgtcaaatgcccaaatcagag--attaccaagaatcgga 1922
Db 1920 GGCACCTCTGGGAGGGCCCTCCAAAGCCACACAAAGGACGATCATCAAGATCGGGA 1979
QY 1923 tgcattgtaagaacaattcagggcattgagcaccctggcggtcactcctggaagcacag 1982
Db 1980 CATCTCTCTCAGGACATCGCCAGTAAAGGAGTTTACAGCTACTTCCAAAGACATCA 2039
QY 1983 cgaagccttgagccctgagagaatgaatcaagaactcccgcggtgagaaactctg 2042
Db 2040 CGAGGTCCTTAACAGAACTGGAAAGGCTACCAAGCTGTAAAGATTTGGAGCGAGTGA 2099
QY 2043 cagagacttgagtcagaaaggtgtgttacctaccgtccacacacttctcctcgcc 2102
Db 2100 CAAGGATTTGAGTTCGAGAGGTCTGCTACTTGTCTCTCAACACGTTTCTGCTGAAGCC 2159
QY 2103 actgcaccggtctatgcatacaagcaggtccttggagcggctgtgtgcaaacaccaccgcc 2162
Db 2160 CATCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2219
QY 2163 gaggcagccgacttcagagactcgcagcgccttggcagagatcagcagagatggtgc 2222
Db 2220 CGGGCCACCATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2279

QY 2223 acagctccacgggtacgatgatcaagaatgagaaattccagaagctgcaagaactcaagaa 2282
Db 2280 CACACTACAGACATCTCTATCCGGCTGGAGAACCTGCAGAACCTACGGAGGTGAGCG 2339
QY 2283 agatttgattggcattgacaactcttggttccgggaaggaggttccatcctgctggcag 2342
Db 2340 GGACCTGGTGGGCATAGAGAACCTCATTTGCTCTCGCAGGAGTTCATCCGTGAGGCTG 2399
QY 2343 cctcagcaagctctcggggaagggttcacagcagcagcatgttctcctgttcaacagct 2402
Db 2400 CCTTCAAGACTCACAAAGAGGGCTGCAGCAGAGGATGTTTTCCTGTTCTCAGATAT 2459
QY 2403 cctgtatatacagcagccggggctgagccctccaatcagtttaaggtccacggcagct 2462
Db 2460 GTTGTCTACACAAAGAGGATTCAGGGACCACCCACTTCGGATCCGGGGCTCCT 2519
QY 2463 cccgtctatggcatgacgattgagagagcgaagcaggtgggggtgccccactgcct 2522
Db 2520 TCCCTTCCAGGCATGCTGTGGAAGAAAGTATAACGAGTGTCTGTTCACACATGTTT 2579
QY 2523 gaccttccggggccagcggcagtcacatcgttggcggccagttctcgttccgagatgga 2582
Db 2580 CACCATCTACGCGGCTCAGAAACAAATCGTGTGGCAGCCAGCACCTCGGTGAGAAAG 2639
QY 2583 gaagtgggttgaggacatccagatggccttgccttgcctggcggagagcagcagccccc 2642
Db 2640 GAAGTGGATGCTGGACTGAATCCCGATCCNAGCAGCCNAGAGTGGCGGTGACACGGC 2699
QY 2643 cctgagttcctggccagca-----gccccctgacaacagctcccctgatgaagccac 2696
Db 2700 CCTGCACTGCCAGGCGGCACCTGTGTGCATCTGTCTCCCGCAGATCCCCAACAGGTATC 2759
QY 2697 cgcgctgaccaggagtcagagagatgacctgagcgcctcgcacacatcgtggagcgca 2756
Db 2760 TCTG---GAGCAGGAGTCAGAAGATGATCTCGGGGTGTCCGCGAGCTCCTTGAGGGGCA 2816
QY 2757 ggcgcgcgcgcgcgcgcacacaaatggtgcagctgtgtggtgacccgcacacacagcgtctc 2816
Db 2817 TGGCCAGCACCGGGCCACACCAATGACAGTGTGCTGTGTACCGGACACACAGCGTGT 2876
QY 2817 catgttgacttcagcagtcagtgaggaatcagttgtgtggaacctcgtgaggaact 2876
Db 2877 CAGGGCAGCACCACTGACGTGCTGAGAACCAAGCTTTTTCAGGATATCTGTAAGAAGTT 2936
QY 2877 caaaacacacagcgggtgcaagaagctggtgtgtgtgttcacaaactctcgtcttctt 2936
Db 2937 CAAAACACTCATGCTGGCAGAGCTCTGGGTGCTTTTACCAACTTCTGTTGTCTT 2996
QY 2937 ctacaaatcacaccaggacaatcatcccttgcagcctgcctcctgctcgtgactcgt 2996
Db 2997 CTACAAAACCTCATCAGGATGACTACCCACTGGCCAGCCCTCCGCTGCTGGCTACAGCGT 3056
QY 2997 caccatccctcctgagtcgagaaacatccagaagaactcgtgttcaagctgcacttcaa 3056
Db 3057 GAGCATCCCCAGGGGGCGGATGGCATACACAAGACTATGTTTTCAGCTCCAGTCAA 3116
QY 3057 gtccacgtctactctcagcggcgaagcagtcacagcttcgaagctgagatgagat 3116
Db 3117 ATCCCACTGCTACTTCTCCGGCTGAGAGCAAGTACACATTTGAAAGGTGGATGAGGT 3176
QY 3117 gatcgcagtcgaccacagctcgtcctcgcgacccc 3151
Db 3177 GATCCAGGGGGCCAGCAGCTCAGCCGGAGGGGCC 3211

RESULT 4

BC009153

LOCUS

DEFINITION

ACCESSION

VERSION

BC009153 3719 bp mRNA ROD 12-JUL-2001
Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304
IMAGE:2655209, mRNA, complete cds.
BC009153
BC009153.1 GI:14318718

KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 3719)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 5 Row: m Column: 20.

FEATURES

Location/Qualifiers
1..3719
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:6304 IMAGE:2655209"
/tissue_type="Mammary tumor; MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."
/clone_lib="NCI_GAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
99..3296
/codon_start=1
/product="Similar to KIAA0793 gene product"
/protein_id="AAH09153.1"
/db_xref="gi:14318719"
/translation="MGEIGTYRALPTSGTRLGQTAIGVSTLEPEQSLSPRMOEKHM RIRKLLDSTVELFDIPKCDGVLLTQVWKHLNIECDYDFGLFKNVQSYWILEPM KPIIROVKPKNAVLRAVKEFPDPGLOEYTRYFLALQKRDLLERLCTTANTA ALLSHLLQSEIGYDETLDRHLKANEYLNQEKSELEKILDFHQRHTGQTPAESDQ VLEIARKLEMTGIRFHMSDREGTKINLAVSHMGVLVPQGTIKINTFNWSKRLSEF KRRLILHPVHGPGYQDTLEFLIGSDECKNFWKICVEYHTFFLSDPKPKAKAVF FSRGSSFRYSGRTQKQVYVYKDGMRKIPYERRHSKTRTSLHALTVDLPLKQSVFTD GLRTSASLSANVFPYPPSSSLPPGPNLKDSSSLVDPOAPVIKSTAERSGSPS SDDGSPSTQSAHLPGPVLVPCGFGFSDSPSPSPSLKSHLSLCPQLQALSTAEQAS PVLSPVLSGATARMQNEQOKHKHPEDEAYFAKEILLATERYLKLDEVTWFRS VLIKEAPGALMALLSNIDPVTEFHGFDUEVEQRUALWEGSPSHLKHGDHGRIGD ILLNRMLQKLETFQRHDEVLELEKATCKKLEAVYKEFELOKVCYLPPLTFLL KPVORLYRLLLSRLCAHYSPGHRDADCEALKAITVETELQOSSLTRLENLQKLT ELQRODLVGLMVESENEVSLHCFTIYAQKTIIVVAASRLERKWMQDLNIAIOA RIRGLPLRGLMVESENEVSLHCFTIYAQKTIIVVAASRLERKWMQDLNIAIOA AKTIGDSPVLLGPGVTRTPRSDEVSLESEDEGRNGRSGLENSQHRANTTHVCHV YRNTSVSRADHSAVENQSLGKYLKFNKSNQWKLWVFTNFCFLFFYFKHDDYPLA SLPLGLVSLPRADSTHKDYFKLQFKNVYFFRAESKYTFERWMDVIRASSSPG RPPSFTQDCSHHSPGLEAIREKEACSPCLDKNL"

CDS

BASE COUNT 997 a 979 c 929 g 814 t
ORIGIN
Query Match 29.6%; Score 1019.2; DB 10; Length 3719;
Best Local Similarity 59.1%; Pred. No. 9.4e-223;
Matches 1867; Conservative 0; Mismatches 1263; Indels 30; Gaps 6;
QY 70 ccgagccaggtacgactggtggggcccgagaaattcgggagtcagtcaccttggaacct 129

Db 132 CCAACTTCAGGACACGCTTGGTGGCCAGACCCGACCTGGAGTGGAGTACCTCGAGCCA 191
QY 130 ggacagagccgcccccaacaccttcaggaatacctgtgtccatcaaaatccagatgctg 189
Db 192 GAGCAGAGCCTGTGTCACCCAGGATGCGAGGAGAACATCGCATCAGAGTGAAGTGCTG 251
QY 190 gatgacacccagggcatttggaattccacaaagagctcctgggaaggtgctggtgat 249
Db 252 GACAGCATGTAGAGTTATTTGACATCGAGCCGAAAGTGTGACGCGAGCTCTTACTAGT 311
QY 250 gcagtttcaacacacctcaacctcgtgaagtgactattttggctcagtttctctgat 309
Db 312 CAAAGTGTGAACATTTGAATCTCATTTGAATGTGACTACTTTGGTCTGAGTTCAGAA 371
QY 310 cacaaaaagatcacggtgtggtggtatctcctaaacaccttgtaaacagattagaag 369
Db 372 GTCCAGTCTCTACTGATTTGGCTTGAACCTATGAACCCATCATTAGGCAAGTACGAA 431
QY 370 ccaagacgctgtgttaagtgtgtaaatcttccgcctgacacacacacacacacac 429
Db 432 CCAAAAAATGCGGTGCTTGCCTGCGCTGAGTAAATTTTCCCGCTGACCTGTCAGCTG 491
QY 430 caagaagaactcacaggtacacctgttcgctgcaggtgaagcagagacttgctcaagc 489
Db 492 CAAGAAGAGTACAAAGGTACCTGTTGCTTGCCTCAACTCAAGAGAGACCTCTCTGG 551
QY 490 aggttgagctgtaatagcaccagcgcgctctcttgatttcacacattgtcaactgtg 549
Db 552 CGCTTGACCTGCGGCCAACACTGCAGCCCTTCTCATATCCACCTTCTGCGAGTCGGA 611
QY 550 attggggtatttgtaagccttgacagagacacttgacagagacacttgacagagaa 609
Db 612 ATCGGAGATTATGATAAACCCTGGATCGAGAACACTCAAAAGCCCAATGAATACTTGC 671
QY 610 cagaagacgactagagagaaataatcggtgaatttcacataaacacattgagaaaca 669
Db 672 AACAGGAGAAATCTCTAGAAAGACTAGACTTCCATCAGAGGCACACGGCCAGACT 731
QY 670 ccagcagatcagatttccagctccttagagattgcccgtcggtgagatgattggaatc 729
Db 732 CTGCGAGAGTCAGATTTCCAGGTGCTTGAGATTGCAAGGAGCTGGAATGATGGCATC 791
QY 730 cgggttgaccccgccagagacaggaagacagagataatctgccccttgcccaacag 789
Db 792 AGGTTTCACTAGCTTCTGACAGAGAGGAGCAAGATTAATCTAGCAGTTTCTCACATG 851
QY 790 ggaattctagtttcagggtttcactaagatcaatgctcactcagctggcgaaggtgcg 849
Db 852 GGTGCTCGGTGTTCCAGGGTACCACCAAAATCAATACCTTCACTGGTCCAAAGTCCGT 911
QY 850 aagctgagcttcaagaggaagcgtttctcactcaagctcgcgcgagatgccaatagtcg 909
Db 912 AAATTAAGCTTCAAGAGAAAAAGGTTCTTATCAAACTCCACCTCGAGTCCATGGGCC 971
QY 910 taccgggataccttggaattctctgattgcccagtcgggatttctgcaagtccttctgaa 969
Db 972 TACCAGGACACGTTAGAGTCTTGTGGTAGCAGAGATGAATGTAAGAACTTCTGGAAG 1031
QY 970 atctgttgaaacatcctcttcttaagacttttgaagagcccaacacacacacacag 1029
Db 1032 ATATGTGTGGATACCATACCTTTTATAGACTCTCTGACAGCTTANGCCAAAGGCAAG 1091
QY 1030 cccgtcctcttagcgggggttcactattcgttgcaggttcggtcagactcagaagcaggt 1089
Db 1092 GCTGTCTTCTCAGCGGAGGCTCCTCTTCAGATACAGTGAAGAACTCAGAAACAATA 1151
QY 1090 ctccgactatgttaagagagagagacataagaaggtgagtttgaaaggagacagcaag 1149
Db 1152 GTAGATTATGTCAAAGACGCTGGAATGAAGAAATTTCCATACGAAAGACGACAGTAAG 1211
QY 1150 attc---attctatccggagccttcttcacagcctcacaagaactgaattcggaggtc 1206

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 2888)
Bloecker, H., Boeche, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2000) MIPS, Am Klopferspitze 18a, D-82152 Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp762P046) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
FEATURES
Location/Qualifiers
1..2888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="DKFZp762P046"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="melanoma (MeWo cell line)"
polya_signal
2852..2857
polya_site
2859
BASE COUNT 879 a 592 c 620 g 797 t
ORIGIN
Query Match 10.0%; Score 342.8; DB 9; Length 2888;
Best Local Similarity 68.2%; Pred. No. 7.5e-68;
Matches 476; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
QY 445 aggtacctgttcgctgcagtgagcaggagctgctcaaggcaggtgacgtgtaaat 504
DB 123 AGGTATCTTTTACTCTCAATAAAGAGGATTTGGCTCTAGGAGGCTTCCATGCAGT 182
QY 505 gacacagcgcagctcttatttcaacattgtgcaactgtgagattggggattttgat 564
DB 183 GACAACTGTACAGCGTGTGATGTCACATCTTACAACTCAGAACTGGAGACTTTTAT 242
QY 565 gaagccttgacagagacacttagcaaaaaataataacataccctcagcaagcagcacta 624
DB 243 GAAGAAACAGATAGGAAGCATCTGGCACAACCTCGGTACTTACCAACCAAGACTGT 302
QY 625 gaggcaaaatcgtggaattccacataaccacattggacaacacaccagcagaatcagat 684
DB 303 GAGGCAAGATCATGCATCTTATCAGAGCAGATTTGGCAGGAGCCACCTGAACTTGAC 362
QY 685 ttccagctcctagagattgcccgtggctgagatgatgtggaatccggttgaccggccc 744
DB 363 ATTCTGCTACTGACATAGCAAGGAGCTGGATATGTATGGCATCAGGCTCACCCCGCC 422
QY 745 aaggacaggaaggaagcagagatcaatctggcgttgcccaacacgggaattctagtgtt 804
DB 423 AGTAGGTTGAAGGGATGATGATTCACCTGGCTGTGCTACATGGGAGTGTGTTA 482
QY 805 caggggttcactaaagatcaatgcttcaactggcgcaaggtgcggaagctgagcttcaag 864
DB 483 CGGGGAATACAAAGATCAATACTTTTACTGGGCTAAATCCGCAAGTTGAGTTTAA 542
QY 865 aggaagcgtttctctcaagctccggccagatgccaatagtgcgtaccagataaccttg 924
DB 543 AGAAAGCATTTTCTCATCAAACTTCATGCCAATATCTTGGTGTGTGGCAAGGATACCT 602
QY 925 gaattcctgatggccagtcgggatttcttcaagtcctctcttgaaaaatctgtgtgaacat 984

DB 603 GAGTTCACCATGGCCAGCCGAGATGCTCAAGGCTTTCTGGAAGACTTGTGTGANTAC 562
QY 985 catgcctcttttagacttttgaagagcccaaacaccccaagccgcctctcttttagc 1044
DB 663 CATGCTTCTTCAGGCTTTCCGGAAGAGCCCAATCAAGCCCAAAACCTACTCTCAGC 722
QY 1045 cgggggtcatctatttcggttcagtgctcggaactcagaagcaggttctcactatgttaa 1104
DB 723 AAGGTTTCAGTTTTCGCTATAGTGGACGAACCCCAAGGCACTTTTGAATATGGAGA 782
QY 1105 gaagagagacataagaagctgaggttgaaggaagca 1142
DB 783 AAAGGAGGCTGAAGAGCTTGCCATTTGAAGGAACA 820
RESULT 6
AL137249
LOCUS AL137249 181800 bp DNA HTG 03-AUG-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-111L24 map q31.3-32.3, ***
SEQUENCING IN PROGRESS ***, 4 unordered pieces.
ACCESSION AL137249
VERSION GI:15131444
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Garner, P.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Aug 9, 2001 this sequence version replaced gi:15131193.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA31K22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 180516 bases at least Q40
Consensus quality: 180899 bases at least Q30
Consensus quality: 181168 bases at least Q20
Insert size: 181500; sum-of-ctigs
Insert size: 174530; 1.8% error; agarose-fp
Quality coverage: 6.26x in Q20 bases; sum-of-ctigs Quality
coverage: 6.72x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 114302: contig of 114302 bp in length
* 114303 114402: gap of 100 bp
* 114403 135098: contig of 20696 bp in length
* 135099 135198: gap of 100 bp
* 135199 173346: contig of 38148 bp in length
* 173347 173446: gap of 100 bp
* 173447 181800: contig of 8354 bp in length.
* Location/Qualifiers
1..181800
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source

Db	128	CGACATGTTGTGCACACTCGTCCCGCCAGATCCCGCCACGAGGATATCTCTG---GACACAGAG	184
Qy	2713	tcagagatgatcctgagcgcctcgcgcacatcgcctggagcgcgcagccgcgcacccgcgc	2772
Db	185	TCAGAAGATGATGCTCGGGGTGTCCGCAGCTCCCTTGAGGGGCGATGGCAGCAGCCGGCC	244
Qy	2773	aacacaatggtgcacgtgctgtgacacgcgcgaacacacgcctccatggtggaacttcagc	2832
Db	245	AACACCACAATGCACGTGTGCTGTGACCGGAACACCGCGTGTCCAGGGCAGACACAGT	304
Qy	2833	atcgagtgagaaatcagttgtctggaacacctgctgaggaataatcaaaaacagcaacgqg	2892
Db	305	GCAGCTGTCGAGAACACAGCTTTCAGGATATCTGCTAAGAAAGTTCAAAACAGTCATGGC	364
Qy	2893	tggcagaagctgtgggtggtgttcacaaacttcctcgtgttcttcttctacaaatcacacag	2952
Db	365	TGGCAGAAGCTCTGGGTGCTCTTTACCAACTTCTGTTTGTCTTCTACAAAATCATCAG	424
Qy	2953	gacaaatcctcctgtcgaagctgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt	3012
Db	425	GATGACTATCCCACTGGCCAGCTCCCGTGTCTGGGCTACAGCGTGAGCATCCCGAGGAG	484
Qy	3013	tcgagaaactccagaaagactacgtgttcgaagctgcacttcaagtcacagtcacgtctactac	3072
Db	485	GCCGATGGCATACACAAAGACTATGTTTTCAGCTTCCAGTTCAAAATCCACAGCTACTTC	544
Qy	3073	ttcaggcggaagcgcagtcacagtttcgaaaagtgtgagtgagtgatccgcagtgccacc	3132
Db	545	TTCCGGGCTCAGAGCAAGTACACATTTGAAAGGTGGATGGAGGTGATCCAGGGGCCAGC	604
Qy	3133	agctcgtcctcgcgcaccc 3151	
Db	605	AGCTACCGGGAGGGGCC 623	
RESULT	9		
AL161896			
LOCUS	AL161896	96183 bp	DNA
DEFINITION	Human DNA sequence from clone RP11-261P24 on chromosome 13, complete sequence.		
ACCESSION	AL161896		20-JAN-2001
VERSION	AL161896.15	GI:12330752	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 96183)		
JOURNAL	Phillimore,B.		
COMMENT	Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
	On Jan 22, 2001 this sequence version replaced gi:12329254.		
	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that a variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences only a small overlap as described above.		
	This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at		
	http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at		

requests: clonerequest@sanger.ac.uk
 On Nov 8, 2000 this sequence version replaced gi:11061750.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 This sequence is the entire insert of clone RP11-10G5 The true left
 end of clone RP11-573N10 is at 109104 in this sequence. The true
 right end of clone RP11-72J7 is at 95284 in this sequence. This
 sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-10G5 is from the
 library RPCI-11.1 constructed by the group of Pieter de Jong. For
 further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6.

FEATURES

source	Location/Qualifiers	misc_feature	misc_feature
1..170916	/organism="Homo sapiens"		/note="63 copies 2 mer ca 88% conserved"
/db_xref="taxon:9606"			41206..41622 GSS: Em:AQ184690"
/chromosome="13"			/note="match: GSS: Em:AQ184690"
/clone="RP11-10G5"			42604..43109
/clone_lib="RPCI-11.1"			/note="match: GSS: Em:AQ573113"
384..873			43382..43429
/note="MER31B repeat: matches 1..481 of consensus"			/note="24 copies 2 mer ta 93% conserved"
1374..1407			complement(46659..47093)
/note="17 copies 2 mer tt 85% conserved"			/note="match: GSS: Em:AQ805998"
5470..5533			46717..46754
/note="32 copies 2 mer tt 70% conserved"			/note="19 copies 2 mer gt 100% conserved"
6352..6714			53170..53274
/note="match: GSS: Em:AQ698072"			/note="MER34 repeat: matches 8..115 of consensus"
8498..8665			53278..53412
/note="TIGGER1 repeat: matches 2080..2238 of consensus"			/note="MER74B repeat: matches 50..180 of consensus"
8670..8888			55996..56438
/note="MER75B repeat: matches 18..239 of consensus"			/note="MER83 repeat: matches 1..448 of consensus"
13335..13686			59536..60314
/note="176 copies 2 mer tg 69% conserved"			/note="MER21B repeat: matches 63..794 of consensus"
13996..14238			60344..60392
/note="L1PB3 repeat: matches 5894..6150 of consensus"			/note="MER21B repeat: matches 741..790 of consensus"
/note="15 copies 2 mer ta 86% conserved"			60825..61336
complement(17881..18076)			/note="MER74B repeat: matches 61..621 of consensus"
/note="match: GSS: Em:AQ776608"			61616..61665
18339..18804			/note="MER61B repeat: matches 377..425 of consensus"
/note="LRF7 repeat: matches 1..448 of consensus"			61667..62339
19891..20105			/note="LRF9 repeat: matches 1..625 of consensus"
/note="MER57-internal repeat: matches 7171..7387 of consensus"			62342..62683
20131			/note="MER61C repeat: matches 1..384 of consensus"
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 260bp by restriction digest data."			63996..64135
29106..29251			/note="MLT2E repeat: matches 1..149 of consensus"
/note="LRF41 repeat: matches 2..160 of consensus"			64486..64768
31944..32589			/note="MLT2E repeat: matches 149..395 of consensus"
/note="LRF12 repeat: matches 1..671 of consensus"			65150..65369
32590..33323			/note="LJMA5A repeat: matches 6037..6091 of consensus"
/note="PTR5 repeat: matches 743..1478 of consensus"			66076..66299
38726..38851			/note="LJMA5A repeat: matches 5827..6047 of consensus"
			67351..67427
			/note="MLT1A2 repeat: matches 173..250 of consensus"
			67428..67487
			/note="30 copies 2 mer ag 88% conserved"
			67492..67660
			/note="MLT1A2 repeat: matches 5..174 of consensus"
			68268..68810
			/note="MER68A repeat: matches 1..561 of consensus"
			complement(73998..74451)
			/note="match: GSS: Em:AQ695161"
			74143..74773
			/note="LTR39 repeat: matches 172..794 of consensus"
			75063..75233
			/note="LTR39 repeat: matches 1..172 of consensus"
			79181..79291
			/note="MLT1A2 repeat: matches 251..374 of consensus"
			79699..80013
			/note="MLT1A2 repeat: matches 1..251 of consensus"
			80095..80463
			/note="THE1C repeat: matches 1..369 of consensus"
			83250..83460
			/note="Tigger3b repeat: matches 1..210 of consensus"
			83757..84298
			/note="Tigger3b repeat: matches 210..750 of consensus"
			84581..84607
			/note="Tigger3b repeat: matches 750..775 of consensus"
			84608..84963
			/note="THE1C repeat: matches 4..371 of consensus"
			84964..85004
			/note="Tigger3b repeat: matches 775..825 of consensus"
			85677..86088
			/note="Tigger3b repeat: matches 823..1231 of consensus"


```
QY 585 cttagcaaaaaataacatacc-----tcagcaagacgcacacagagacacaaatcgt 638
Db 1044 TTATGTTAGTGAATTTTAAACTGGCCCGGAATCAGACCAAGGAAGTCAAGAGGTCAT 1103
QY 639 ggaattccacataacacacattggacacacacacagacagatccagattccagctccataga 698
Db 1104 GGAACCTGCAATAGTCATACAGGTCCATGACTCCAGCTCAGGCTGACTTGGAGTTCTTGA 1163
QY 699 gattgcgcgtcggtagagatgatgaatccggtttcaccccgccacagacagaggaag 758
Db 1164 GAATGCCAAAAAGTTGCTATGATGAGTTGATCTTCATACCAAGCAAGGACTTGGGAAG 1223
QY 759 cagcaagatcaatctggccgttgcacacagcgggaattctagtgttccaggtttccataa 818
Db 1224 AGTAGATATCATCTCAGTGTCTGCTCTAGTGGCTTCTGCTTTTACAAAGATACAGCTGAG 1283
QY 819 gatcaatgcctcaactggccaggtgcgaagctgagcttcaagaggaagcgctttct 878
Db 1284 AATTAACCGTTCCTTGGCCCAAGTGCCTGAGATTTCTTATACAGTATGACTTTT 1343
QY 879 cataagctcgcgcagatgccaatagtcgctaccagatcaccttgaattctctgatggc 938
Db 1344 CATCAAGATTCCGGCTCGAGAGCAAGCAGATGAAAGTACCATCGGATTCAAACTTCC 1403
QY 939 cagtcgggattctgcagctcctctggaataatctgtgtgaacatcatcctctcttag 998
Db 1404 CAGTTACCGAGCAGCTAAGAAATTTATGAAAGTCTGTGTAAGATCACACAGTTCATCG 1463
QY 999 acttttgaagagcccaacacacagccagccgctctcttagccggggtcatcatt 1058
Db 1464 A---TTGACATCTACAGACACATTCCTCCAAAGCAAAATTTCTTCGCTAGATCCAAAT 1520
QY 1059 tcggttcagtggtcgactcagaagcag 1086
Db 1521 TCGATACAGTGGCGGACTCAAGCTCAG 1548

RESULT 13
AC022669
LOCUS Homo sapiens chromosome 15 clone RP11-96B23 map 15, WORKING DRAFT
DEFINITION AC022669 152053 bp DNA HPG 04-MAY-2001
SEQUENCE 33 unordered pieces.
AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HPG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhaltier,B., Brown,A., Burkett,G., Castle,A.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
McDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
MacPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severi,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:7229803.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4938
Center clone name: 96_B_23
----- Summary Statistics
Sequencing vector: M13; M77815; 93% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136451 bases at least Q40
Consensus quality: 143612 bases at least Q30
Consensus quality: 146365 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 148853; sum-of-contigs
Quality coverage: 3.4 in Q20 bases; agarose-fp
Quality cov.
NOTE: This is a 'working draft' sequence. It currently
consists of 33 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 784: contig of 784 bp in length
785 884: gap of 100 bp
885 1956: contig of 1072 bp in length
1957 2056: gap of 100 bp
2057 3113: contig of 1057 bp in length
3114 3213: gap of 100 bp
3214 4817: contig of 1604 bp in length
4818 4917: gap of 100 bp
4918 6582: contig of 1665 bp in length
6583 6682: gap of 100 bp
6683 8059: contig of 1377 bp in length
8060 8159: gap of 100 bp
8160 22369: contig of 14210 bp in length
22370 22469: gap of 100 bp
22470 24974: contig of 2505 bp in length
24975 25074: gap of 100 bp
25075 27206: contig of 2132 bp in length
27207 27306: gap of 100 bp
27307 29420: contig of 2114 bp in length
29421 29520: gap of 100 bp
29521 32000: contig of 2480 bp in length
32001 32100: gap of 100 bp
32101 36330: contig of 4230 bp in length
36331 36430: gap of 100 bp
36431 39067: contig of 2637 bp in length
39068 39167: gap of 100 bp
39168 42564: contig of 3397 bp in length
42565 42664: gap of 100 bp
42665 46664: contig of 4000 bp in length
46665 46764: gap of 100 bp
46765 51047: contig of 4283 bp in length
51048 51147: gap of 100 bp
51148 52569: contig of 4122 bp in length
52570 55369: gap of 100 bp
55370 58831: contig of 3462 bp in length
58832 58931: gap of 100 bp
58932 64077: contig of 5146 bp in length
64078 64177: gap of 100 bp
64178 68808: contig of 4631 bp in length

Query Match 5.6%; Score 193.4; DB 2; Length 152053;
Best Local Similarity 97.0%; Pred. No. 9.3e-34;
Matches 197; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1508 cgcaggggggagtgggccctcgccacgtgaccttgcctcccaacctgagcccgacacca 1567
 101390 cgcaggggggagtgggccctcgccacgtgaccttgcctcccaacctgagcccgacacca 1567
 Db 101390 CGAGGGGGGAGTGGCCCTCGCCACGTGACCTTGCTCCCAACCTGAGCCCGACACCA 101449

Qy	1568	agaagccttccttgatcaagcccgctgtgtaataacacagagcctgcccccagacgacg	1627
Db	101450	AGCAGGCCTTCCTCTTGATCAGCCCCGCTGCTGAATGACGAGGCTGCCCCCGAGGAGC	
Qy	1628	atgagatgagggccgaggaag	1650
Db	101510	ATGAGGATGAGGGCCGGAGGAG	101532

RESULT	14	
LOCUS	HUMELIA	
DEFINITION	Human structural protein 4.1 mRNA	27-APR-1993
ACCESSION	M14993	
VERSION	M14993.1 GI:182075	
KEYWORDS	protein 4.1.	
SOURCE	Human reticulocyte, cDNA to mRNA, clones lambda-HE 4.1-8, lambda-HE 4.1-6, and lambda-HE 4.1-A.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 2867)	
TITLE	Conboy, J., Kan, Y. W., Shohet, S. B. and Mohandas, N.	
JOURNAL	Molecular cloning of protein 4.1, a major structural element of the human erythrocyte membrane skeleton	
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 83, 9512-9516 (1986)	
REFERENCE	87092279	
AUTHORS	2 (sites)	
TITLE	Conboy, J., Marchesi, S., Kim, R., Agre, P. A., Kan, Y. W. and Mohandas, N.	
JOURNAL	Molecular analysis of insertion/deletion mutations in protein 4.1 in elliptocytosis; Determination of Molecular genetic origins of rearrangements	
MEDLINE	J. Clin. Invest. 86, 524-530 (1990)	
COMMENT	90347048	
FEATURES	[2] sites; insertions and duplications causing elliptocytosis.	
source	Location/Qualifiers	
	1..2867	
	/organism="Homo sapiens"	
	/db xref="taxon:9606"	

Qy	819	gatacatccttcaactcggcccaaggctgcggaaagctcagcttcacaggaggagcgtcttct	878					
Db	1455	AAATTAAACGGTTCCTCTTGGCCCAAGTCTGCTGAAGATTTCTTATTAAGCGTAGTACGCTTTT	1514					
Qy	879	catcaagctcggcgagatgccaatagtcggtaccagagataccttgaattcctcgatgagc	938					
Db	1515	CATCAAGATTTCGGCTCGAGAGCAAGAGCAGATATGAAGAATACCATCGGATTCAAACTCC	1574					
Qy	939	cagtcgggatttctgcaagtcctcttgaaatactgtgtgaacatcatcgctctctttag	998					
Db	1575	CAGTTACCGAGCAGCTAAGAAATTTATGAAAGTCTGTGTAGAACATCACACGTTTTTTCAG	1634					
Qy	999	acttttgaagagcccaaccccaagcccaagccgctcctctttagcggggggtcatcatt	1058					
Db	1635	A--TTGACATCTACAGACACCAATTCGCCAAAGCAAAATTTCTTCGCTAGGATCCAAAT	1691					
Qy	1059	tcggttcagtggtcgagactcagaagcag	1086					
Db	1692	TCGATACAGTGGCCGACTCAAGTCAG	1719					
RESULT 15								
HUMEMP41								
LOCUS	HUMEMP41	3043 bp	mRNA PRI 03-MAR-1994					
DEFINITION	Homo sapiens erythroid membrane protein 4.1 mRNA, complete cds.							
ACCESSION	M61733							
VERSION	M61733.1 GI:182081							
KEYWORDS	erythroid membrane protein 4.1.							
SOURCE	Homo sapiens cDNA to mRNA.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS	Conboy, J.G., Chan, J.Y.C., Chasis, J.A., Kan, Y.W. and Mohandas, N.							
TITLE	Tissue- and development-specific alternative RNA splicing regulates expression of multiple isoforms of Erythroid membrane protein 4.1							
JOURNAL	J. Biol. Chem. 266, 8273-8280 (1991)							
MEDLINE	91217063							
FEATURES	Location/Qualifiers							
source	1..3043							
CDS	/organism="Homo sapiens"							
	/db_xref="taxon:9606"							
	816..2741							
	/codon_start=1							
	/product="erythroid membrane protein 4.1"							
	/protein_id="AAA35797.1"							
BASE COUNT	/db_xref="GI:182082"							
	/translation="MCKVSVLLDDTVYECVVEKHAKGQDILLKRVCEHLNLEEDYFGLA AINDNATSKTWLDSAKEIKKQKGVGNWFTFNKFFPPDPAQFETIRIYLCQLQKRA DIVAGRLPCPTAQADLELGSYIQSELGDYDELHVDVDFDLAPNQTKELEKNWEM LHKYSRMSFQADLLEFNALKLSMYGVDLHRAKDLGVDIILGVSSGILLYVKDLK RINRPMWPKVLKYSKRSSFYKIRPGEQOYEISTGFKLPSYRAAKKLWKVCVRSHT FRLTSDTITPKSKFLALGSKFSYSGTQATQOASALIDRPAPHFRTAKSRASRIS DGAAAVDSADRSFPRSPAPAITQGVAAEGGVLDASAKKTVVPAKQETVAEYKKEDEK PPQAEPEPTAEAWKVEHLIEVTPTSNGDQTKRRLDGENIYIHSNLMLEDDLK SQEIEKKHNASISELAKNFMSEVPEPRSEWKRSLSPFTLNINQGITGEGPFL VKTQTVTISDNANAKSEIPTKDVPIVHTETKITVEAAQTDONSQGLDQGVLLTAQT ITRETSSSTTTTOITTKTVKGGISETRIEKRIKIVITGDADIDHDQVLQVAKAEKQHDH MSYTKVHVHVEETRIADE"							
	997 a	673 c	690 g					
	583 t							
	BASE COUNT							
	ORIGIN							
Query Match 5.68; Score 192.4; DB 9; Length 3043;								
Best Local Similarity 52.88; Pred. No. 2.le-33;								
Matches 490; Conservative 0; Mismatches 426; Indels 12; Gaps 3;								
Qy	165	cgtgtccatcaaaatcccatgctggtgagacacccagagcatttgaagttccacaaag	224					
Db	815	CATGCATGCAGGTTCCTTTGTTGGATGACACAGATTTATGAATGCTGTTGGAGAAACA	874					
Qy	225	agctcctgggaaggtgctgctggatgcagtttgcacaccacatcctcgctgggaagtg	284					

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:55:11 ; Search time 36.72 Seconds
(without alignments)
55.768 Million cell updates/sec

Title: US-09-555-342A-2_COPY_764_854

Perfect score: 468
Sequence: 1 GSKLSKGLQRMFFLN.....IVAASRSEMEKWKVEDIQMA 91

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.5	18.1	394	2	US-08-839-581A-4
2	84.5	18.1	394	4	US-09-023-591A-4
3	80.5	17.2	814	1	US-08-318-831-2
4	79.5	17.0	1244	3	US-09-356-952-7
5	75.5	16.1	399	2	US-08-839-581A-2
6	75.5	16.1	399	4	US-09-023-591A-2
7	71.5	15.3	397	4	US-09-284-033-4
8	71.5	15.3	397	4	US-08-729-834B-4
9	71.5	15.2	127	4	US-09-284-033-6
10	71.5	15.2	127	4	US-08-729-834B-6
11	71.5	15.2	398	2	US-08-839-581A-32
12	71.5	15.2	398	4	US-09-023-591A-32
13	70.5	15.1	126	4	US-09-284-033-5
14	70.5	15.1	126	4	US-08-729-834B-5
15	70.5	15.1	399	4	US-09-284-033-2
16	70.5	15.1	399	4	US-08-729-834B-2
17	69.5	14.9	537	2	US-08-633-879C-2
18	67.5	14.4	689	1	US-08-221-817-18
19	67.5	14.4	689	1	US-08-454-439-18
20	67.5	14.4	689	5	PCT-US94-10487-18
21	66.5	14.2	136	2	US-08-943-208-2
22	63	13.5	1297	6	US-08-290-731C-4
23	61	13.0	3077	6	5223423-2
24	60	12.8	496	2	US-08-463-418-2
25	59.5	12.7	363	4	US-08-981-234B-2
26	59.5	12.7	527	2	US-08-592-126-145
27	59.5	12.7	527	2	US-08-687-080-48

28	59.5	12.7	535	2	US-08-633-879C-4	Sequence 4, Appli
29	58.5	12.5	668	4	US-09-436-605-2	Sequence 2, Appli
30	58.5	12.5	688	1	US-08-221-817-19	Sequence 18, Appl
31	58.5	12.5	688	1	US-08-454-439-19	Sequence 19, Appl
32	58.5	12.5	688	5	PCT-US94-10487-19	Sequence 19, Appl
33	57.5	12.3	266	3	US-08-815-809-5	Sequence 5, Appli
34	57.5	12.3	280	2	US-08-816-155B-43	Sequence 43, Appl
35	57.5	12.3	280	4	US-09-079-587-43	Sequence 43, Appl
36	57.5	12.3	374	3	US-08-878-801-2	Sequence 2, Appli
37	57.5	12.3	502	1	US-08-484-840-3	Sequence 3, Appli
38	57.5	12.3	502	1	US-08-483-094-3	Sequence 3, Appli
39	57.5	12.3	646	4	US-09-232-200-59	Sequence 59, Appl
40	57.5	12.3	646	4	US-09-232-197-59	Sequence 59, Appl
41	57	12.2	298	2	US-08-838-543-5	Sequence 5, Appli
42	57	12.2	796	2	US-08-817-900-2	Sequence 2, Appli
43	57	12.2	796	4	US-09-236-645-2	Sequence 2, Appli
44	56.5	12.1	659	1	US-08-136-277-3	Sequence 3, Appli
45	56.5	12.1	659	2	US-08-479-403-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-839-581A-4
; Sequence 4, Application US/08839581A
; Patent No. 5958705
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald E.
; APPLICANT: Lipsky, Brian P.
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
; TITLE OF INVENTION: Binding/Signaling
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,581A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/33886
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-839-581A-4

Query Match 18.1%; Score 84.5; DB 2; Length 394;
Best Local Similarity 25.7%; Pred. No. 0.0037;
Matches 29; Conservative 22; Mismatches 33; Indels 29; Gaps. 5;
QY 1 GSKLSKSG--KGLQRMFFLNVDVLLYTSRGLTASNQFKVHGOLPLGYGMTIESEDEMGV 58
DB 264 GWLLKLGKRVKTRWRNFILTDNCLY----FEFTTKPRGIIPLNLSVQKVDDP-KK 318

```

/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Rhone-Poulenc Rorer Inc.
/ STREET: 500 Arcola Road, 3C43
/ CITY: Collegeville
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19426
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: Macintosh
/ OPERATING SYSTEM: System 7.1
/ SOFTWARE: Word 5.1 (EPO Patentin)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/318,831
/ FILING DATE: 19 October 1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: FR92/04827
/ FILING DATE: 21-APR-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Smith, Julie K.
/ REGISTRATION NUMBER: P-38,619
/ REFERENCE/DOCKET NUMBER: ST92033-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (610) 454-3839
/ TELEFAX: (610) 454-3808
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 814 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-318-831-2

Query Match 17.28; Score 80.5; DB 1; Length 814;
Best Local Similarity 26.18; Pred. No. 0.035; Indels 31; Gaps
Matches 29; Conservative 13; Mismatches 38;

QY 1 GSLSKLKGKGLQRMFFLFNDVLVYTSRGLFASNQKFKVHGOLPLYGMTI-----EESDEW 56
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 34 GRGLSLSLKKEGRQCFLFSKHLIICITRGSGKHLTKNGVISLIDCTLLPEPESTEE- 92
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 57 GVPCHLTILRGQRQ-----SIIVAASSRSEMEKQWEDI 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 93 -----AKGSGQIDHLDLFKIGVEPKDSPPTVILVASSRQKAATSDI 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-356-952-7
/ Sequence 7, Application US/09356952
/ Patent No. 6117663
/ GENERAL INFORMATION:
/ APPLICANT: Boriack-Sjodin, Ann
/ APPLICANT: Margarit, S. M.
/ APPLICANT: Bor-Sogli, Dafna
/ APPLICANT: Cole, Philip
/ APPLICANT: Kuriyan, John
/ TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
/ TITLE OF INVENTION: THEREOF
/ FILE REFERENCE: 600-1-228N
/ CURRENT APPLICATION NUMBER: US/09/356,952
/ CURRENT FILING DATE: 1999-07-19
/ EARLIER APPLICATION NUMBER: 60/093,631
/ EARLIER FILING DATE: 1998-07-21
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 1244
/ TYPE: PRT
/ ORGANISM: Rattus rattus
/ US-09-356-952-7

```

RESULT 6
US-09-023-591A-2
; Sequence 2, Application US/09023591A
; Patent No. 6210914

```

RESULT      7
US-09-284-033-4
; Sequence 4, Application US/09284033
; Patent No. 6194173
; GENERAL INFORMATION:
; APPLICANT: Czech, Michael P. and Klarlund, Jes K.
; TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHONOSITIDES, GR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/284,033
FILING DATE: 1999-04-06
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN: 08/729,834
FILING DATE: 07 OCTOBER 1996
APPLICATION NUMBER: PCT/US97/18152
FILING DATE: 1997-10-07
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: UMM-018CPUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-09-284-033-4

Query Match 15.3%; Score 71.5; DB 4; Length 397;
Best Local Similarity 24.3%; Pred. No. 0.21;
Matches 28; Conservative 17; Mismatches 41; Indels 29; Gaps 4;
QY 1 GSLSKLSG---KGLQRMFLFNDVLYTSRGLTASNQFKVHGOLPLYGMTIESEDEWG 57
Db 265 GWLLKGGGRVKTWRWFILTDNCLY-----FEYTTDKPRGIIPLENLSIREVEDS-K 319
QY 58 VPHCLTL-----RGQROSIIVAASSRSEMEKWKVEDIQMA 91
Db 320 KPNCFELYIPDNKDQVIKACKTEADGRWGNHTVYRISAPTPEKEEWIKCIAA 374

RESULT 8
US-08-729-834B-4
Sequence 4, Application US/08729834B
Patent No. 6221841
GENERAL INFORMATION:
APPLICANT: Czech, Michael P.
APPLICANT: Klarlund, Jes K.
TITLE OF INVENTION: General Receptors for Phosphoinositides
TITLE OF INVENTION: and Uses Related Thereto
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,834B
FILING DATE: October 7, 1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: UMM-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-729-834B-4

Query Match 15.3%; Score 71.5; DB 4; Length 397;
Best Local Similarity 24.3%; Pred. No. 0.21;
Matches 28; Conservative 17; Mismatches 41; Indels 29; Gaps 4;
QY 1 GSLSKLSG---KGLQRMFLFNDVLYTSRGLTASNQFKVHGOLPLYGMTIESEDEWG 57
Db 265 GWLLKGGGRVKTWRWFILTDNCLY-----FEYTTDKPRGIIPLENLSIREVEDS-K 319
QY 58 VPHCLTL-----RGQROSIIVAASSRSEMEKWKVEDIQMA 91
Db 320 KPNCFELYIPDNKDQVIKACKTEADGRWGNHTVYRISAPTPEKEEWIKCIAA 374

RESULT 9
US-09-284-033-6
Sequence 6, Application US/09284033
Patent No. 6194173
GENERAL INFORMATION:
APPLICANT: Czech, Michael P. and Klarlund, Jes K.
TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHOINOSITIDES, GRP1 OR GENERAL REC
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/284,033
FILING DATE: 1999-04-06
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN: 08/729,834
FILING DATE: 07 OCTOBER 1996
APPLICATION NUMBER: PCT/US97/18152
FILING DATE: 1997-10-07
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: UMM-018CPUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-09-284-033-6

Query Match 15.2%; Score 71; DB 4; Length 127;
Best Local Similarity 24.1%; Pred. No. 0.051;
Matches 28; Conservative 17; Mismatches 41; Indels 30; Gaps 4;


```
/
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/839,581
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Young J. Suh
/ REGISTRATION NUMBER: P-41,337
/ REFERENCE/DOCKET NUMBER: 27866/33886
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ INFORMATION FOR SEQ ID NO: 32:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 398 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ US-09-023-591A-32

Query Match 15.2%; Score 71; DB 4; Length 398;
Best Local Similarity 24.1%; Pred. No. 0.24;
Matches 28; Conservative 17; Mismatches 41; Indels 30; Gaps 4;

QY 1 GSLSKLSG---KGLQRMFFLNDVLYTSRGLTASNQFKVHGQPLVGYMTIESEDEWG 57
| | | | | : | | | | | : | | | | | : | | | | |
Db 265 GWLLKLGGRVKTKRRWFILTDNCLY---FEYTTDKPRGIIPLENLSIREVEDS-K 319
| | | | | : | | | | | : | | | | | : | | | | |
QY 58 VPCHLTL-----RQQRQSIIVAASSRSEMEKWKVEDIQ 91
| | | | | : | | | | | : | | | | | : | | | | |
Db 320 KPNCFELYIPDNKDQVICKACKTEADGRVVEGNHTVYRISAPTEPEKEEWIKCIKAA 375
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 13
US-09-284-033-5
; Sequence 5, Application US/09284033
; Patent No. 6194173
; GENERAL INFORMATION:
; APPLICANT: Czech, Michael P. and Klarlund, Jes K.
; TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHOINOSITIDES; GRPI OR GENERAL RECEP
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION NUMBER: US/09/284,033
; FILING DATE: 1999-04-06
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN: 08/729,834
; FILING DATE: 07 OCTOBER 1996
; APPLICATION NUMBER: PCT/US97/18152
; FILING DATE: 1997-10-07
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: UMM-018CPUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)742-4214
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
```

```
/
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ FRAGMENT TYPE: internal
/ US-09-284-033-5

Query Match 15.1%; Score 70.5; DB 4; Length 126;
Best Local Similarity 24.8%; Pred. No. 0.059;
Matches 28; Conservative 16; Mismatches 40; Indels 29; Gaps 4;

QY 1 GSLSKLSG--KGLQRMFFLNDVLYTSRGLTASNQFKVHGQPLVGYMTIESEDEWG 58
| | | | | : | | | | | : | | | | | : | | | | |
Db 11 GWLLKLGGRVKTKRRWFILTDNCLY---FEYTTDKPRGIIPLENLSIREVEDP-RK 65
| | | | | : | | | | | : | | | | | : | | | | |
QY 59 PHCLTL-----RQQRQSIIVAASSRSEMEKWKVEDIQ 89
| | | | | : | | | | | : | | | | | : | | | | |
Db 66 PNCFELYNPESHKGQVICKACKTEADGRVVEGNHVYRISAPSPKEKEEMKSIK 118
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 14
US-08-729-834B-5
; Sequence 5, Application US/08729834B
; Patent No. 6221841
; GENERAL INFORMATION:
; APPLICANT: Czech, Michael P.
; APPLICANT: Klarlund, Jes K.
; TITLE OF INVENTION: General Receptors for Phosphoinositides
; TITLE OF INVENTION: and Uses Related Thereto
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/729,834B
; FILING DATE: October 7, 1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: UMM-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-729-834B-5

Query Match 15.1%; Score 70.5; DB 4; Length 126;
Best Local Similarity 24.8%; Pred. No. 0.059;
Matches 28; Conservative 16; Mismatches 40; Indels 29; Gaps 4;

QY 1 GSLSKLSG--KGLQRMFFLNDVLYTSRGLTASNQFKVHGQPLVGYMTIESEDEWG 58
| | | | | : | | | | | : | | | | | : | | | | |
Db 11 GWLLKLGGRVKTKRRWFILTDNCLY---FEYTTDKPRGIIPLENLSIREVEDP-RK 65
| | | | | : | | | | | : | | | | | : | | | | |
QY 59 PHCLTL-----RQQRQSIIVAASSRSEMEKWKVEDIQ 89
| | | | | : | | | | | : | | | | | : | | | | |
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:52:39 ; Search time 44.17 Seconds
(without alignments)
156.936 Million cell updates/sec

Title: US-09-555-342A-2_COPY_764_854
Perfect score: 468
Sequence: 1 GSLSKLSGKGLQORFFLEFN.....IVAASSRSEMEKWKVEDIQMA 91
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	1045	2 JC5795	CDEP protein - hum
2	109.5	23.4	254	2 S40944	hypothetical prote
3	107	22.9	961	2 A55380	facio-genital dyspl
4	102	21.8	316	2 T13601	hypothetical prote
5	99.5	21.3	926	2 T15683	hypothetical prote
6	94.5	20.2	414	2 T22498	hypothetical prote
7	87	18.6	646	2 JC5583	85K SH3 domain-con
8	81	17.3	211	2 T17228	hypothetical prote
9	79.5	17.0	1244	2 S29083	guanine-nucleotide
10	79	16.9	994	1 JQ0151	myosin heavy chain
11	78.5	16.8	875	2 T19678	hypothetical prote
12	77	16.5	496	2 T46356	hypothetical prote
13	76.5	16.3	1275	2 A38985	nucleotide exchange
14	75	16.0	685	2 JC6331	rho-type guanine e
15	74	15.8	807	2 A34581	oxysterol-binding
16	74	15.8	809	2 A34404	oxysterol-binding
17	73.5	15.7	1557	2 T13160	protein CNK - frui
18	72.5	15.5	1260	2 S28407	guanine nucleotide
19	71	15.2	398	2 S24168	SEC7 homolog - hum
20	71	15.2	753	2 S48267	probable membrane
21	70.5	15.1	1737	2 A59235	unconventional myo
22	69.5	14.9	537	2 T49135	prolyl 4-hydroxyla
23	68	14.5	381	2 T46482	hypothetical prote
24	67.5	14.4	189	2 F82448	conserved hypothet
25	67.5	14.4	689	1 A53791	beta-adrenergic-re
26	66.5	14.2	380	2 C75016	na+/h+ antiporter
27	66.5	14.2	689	1 A40088	beta-adrenergic-re
28	66	14.1	252	2 E86104	ATP-binding compon
29	65.5	14.0	273	2 A86672	LysR family trans

30	65.5	14.0	409	2 D64216	hypothetical prote
31	65	13.9	648	2 T33339	hypothetical prote
32	64.5	13.8	374	2 JC7091	centaurin alpha 1
33	64.5	13.8	442	2 E71882	is606 transposase
34	64.5	13.8	570	2 T32743	hypothetical prote
35	64	13.7	230	2 T34541	hypothetical prote
36	63.5	13.6	488	2 T21596	hypothetical prote
37	63.5	13.6	566	2 T16375	hypothetical prote
38	63	13.5	1297	2 S25714	son-of-sevenless-2
39	62.5	13.4	359	2 F83850	hypothetical prote
40	62	13.2	168	2 T40346	omp16 protein - Br
41	61.5	13.1	1166	2 S33812	myosin-like protei
42	61	13.0	275	2 E71516	probable diaminopi
43	61	13.0	513	2 S75997	asparagine-tRNA 1
44	61	13.0	611	1 A55888	protein kinase (EC
45	61	13.0	1115	2 T31068	N-methyl-D-asparta

ALIGNMENTS

RESULT 1
JC5795
C:Species: Homo sapiens (man)
C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C:Accession: JC5795
R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.
Biochem. Biophys. Res. Commun. 241, 369-375, 1997
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain
ing factors.
A:Reference number: JC5795; MUID:98086358
A:Accession: JC5795
A:Molecule type: mRNA
A:Residues: 1-1045 <KOY>
A:Cross-references: DBJ:AB008430
C:Comment: This protein is involved in the adhesion, proliferation, and differentiati
on.
F:1-374/Domain: ezrin-like #status predicted <EZR>
F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>
F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 100.0%; Score 468; DB 2; Length 1045;
Best Local Similarity 100.0%; Pred. No. 4.3e-45;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GSLSKLSGKGLQORFFLEFNVDVLLYTSGLTASNOFKVHGQLPLXGMTIESEDEWGVPH 60
Db 764 GSLSKLSGKGLQORFFLEFNVDVLLYTSGLTASNOFKVHGQLPLXGMTIESEDEWGVPH 823
Qy 61 CLTLRGQRQSIIVAASSRSEMEKWKVEDIQMA 91
Db 824 CLTLRGQRQSIIVAASSRSEMEKWKVEDIQMA 854

RESULT 2
S40944
C:Species: Caenorhabditis elegans
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
C:Accession: S40944
R:Berkas, M.
submitted to the EMBL Data Library, February 1993
A:Reference number: S40933
A:Accession: S40944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <BER>
A:Cross-references: EMBL:Z22181
C:Genetics:
A:Introns: 6/1; 137/1; 206/3; 241/3

```

Query Match          23.4%; Score 109.5; DB 2; Length 254;
Best Local Similarity 30.3%; Pred. No. 6.1e-05;
Matches 27; Conservative 21; Mismatches 40; Indels 1; Gaps 1;

QY 1 GSLSKLSKGGLQORMFELFNDVLATSRGLTASNQFKVHGOLPLGYGMTIESEDEMGVPH 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 GVLVMCKRKPQKQFFLFNDILYI-GNIVTSKKRYNKQRIILRGVQVDELDGDIKKH 98
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 CLTLRGQRQSIIVAASSRSEMEKWVEDI 89
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99 GWIIKTPAKSPAVYAATETEKRENNLHIE 127
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
A53380
C:Species: Homo sapiens (man)
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
C:Accession: A53380
R:Pastorius, N.G.; Cadle, A.; Logie, L.J.; Porteous, M.E.M.; Schwartz, C.E.; Steele, S.;
Cell 79, 669-678, 1994
A:Title: Isolation and characterization of the facio-genital dysplasia (Aarskog-
A:Reference number: A53380; MUID:95042764
A:Accession: A53380
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-961 <PAS>
A:Cross-references: GB:U11690; NID:g595424; PID:g595425
C:Superfamily: CDC24 homology; plectstrin repeat homology
F:373-561/Domain: CDC24 homology <CD24>

Query Match          22.9%; Score 107; DB 2; Length 961;
Best Local Similarity 33.0%; Pred. No. 0.00054;
Matches 30; Conservative 17; Mismatches 40; Indels 4; Gaps 3;

QY 1 GSLSKLSKGK--LQORMFELFNDVLVLTSTRGLTASNQ-FKVHGOLPLGYGMTIESEDEMG 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 595 GHILKLSAKNGTTQDRVLIILFNDRLLYCVPLRLILGQKFSVRARIDVDGMELKESSN-LN 653
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 VPHCLTLRGQRQSIIVAASSRSEMEKWVEDI 88
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 654 LPRTFLVSGKQKRSLELQARTEERKDWQAI 684
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
T13601
Hypothetical protein 80H7.5 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13601
R:Benos, P.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17667
A:Accession: T13601
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-316 <BEN>
A:Cross-references: EMBL:AL031027; NID:el313443; PID:el310143; PIDN:CAA19842.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0000481
A:Note: EG:80H7.5

```

[illegible]

```

RESULT 9
guanine-nucleotide-releasing protein - rat
S29083
N;Alternate names: CDC25 protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: S29083
R;Shou, C.; Farnsworth, C.L.; Neel, B.G.; Feig, L.A.
Nature 358, 351-354, 1992
A;Title: Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor fo
A;Reference number: S29083; MUID:92350260
A;Accession: S29083
A;Molecule type: mRNA
A;Residues: 1-1244 <SHO>
A;Cross-references: EMBL:X67241; NID:g57664; PIDN:CAA47666.1; PID:g57665
A;Note: the authors translated the codon GAG for residues 135 and 137 as Gln
C;Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol
F;240-426/Domain: CDC24 homology <CD24>
F;1005-1241/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 17.08; Score 79.5; DB 2; Length 1244;
Best Local Similarity 26.9%; Pred. No. 0.98;
Matches 29; Conservative 14; Mismatches 38; Indels 27; Gaps 3;

Qy 1 GLSLSKSGKLQQRMEFLFNDVLLYTSRGLTFASNQFKVHGOLPLYGMTI---ESEEDEWG 57
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 GRGSLSKKEGEKQCFKFKHLICTRGSGSKHLTKNGVISLIDCTLLDDPENDDG 536
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 58 VPHCLTLRGQRQ-----SIIVAASRSRSEMEKWVEDI 88
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 -----KGQEVHDLFKINWPKDPPFTVILVASSRQEKAAWTSDI 577
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
JQ0151
myosin heavy chain IA - slime mold (Dictyostelium sp.)
N;Alternate names: ambA protein
N;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Dictyostelium sp.
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Feb-2001
C;Accession: JQ0151
R;Titus, M.A.; Warrick, H.M.; Spudich, J.A.
Cell Regul. 1, 55-63, 1989
A;Title: Multiple actin-based motor genes in Dictyostelium.
A;Reference number: JQ0151; MUID:92096486
A;Accession: JQ0151
A;Molecule type: DNA
A;Residues: 1-994 <TIT>
A;Cross-references: GB:S73909; NID:g241367; PIDN:AAB20711.1; PID:g241368
A;Experimental source: strain A5-2
C;Genetics:
A;Gene: ambA
A;Introns: 1/2; 42/3
C;Superfamily: brush border myosin heavy chain I; myosin motor domain homology
C;Keywords: actin binding; hydrolase; nucleotide binding; P-loop
F;15-706/Domain: myosin motor domain homology <MMOT>
F;105-112/Region: nucleotide-binding motif A (P-loop)
F;594-615/Region: actin binding #status predicted
F;783-994/Domain: actin binding
F;829-935/Region: basic
F;111/Binding site: ATP (Lys) #status predicted

```

```

F;783-994/Domain: carboxyl-terminal <CTD>
F;829-935/Region: basic
F;111/Binding site: ATP (Lys) #status predicted

Query Match          16.9%; Score 79; DB 1; Length 994;
Best Local Similarity 24.5%; Pred. No. 0.87;
Matches 25; Conservative 24; Mismatches 31; Indels 22; Gaps 5;

Qy      3 LSKLSGKGLQORMFFLENDVLLYTSRG--LTASNQFKVHGOLPLYG--MTIESEDEWG 57
      |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
Db      20 LTEVSESSLHENLIRKEGLIYTSIGPVLVSMNPYK---OLGIYGNQDNLNYGKRHEFE 76
      |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
Qy      58 VP-----HCLTLRGQRQSIIVAASS---RSEMEKRW 85

```

C;Accession: A38985; A46199; I58371

[illegible]

RESULT 15
 A34581
 oxysterol-binding protein - human
 C:Species: Homo sapiens (man)
 C:Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 05-Nov-1999
 C:Accession: A34581
 R:Levanon, D.; Hsieh, C.L.; Francke, U.; Dawson, P.A.; Ridgway, N.D.; Brown, M.S.; Golds
 Genomics 7, 65-74, 1990
 A:Title: cDNA cloning of human oxysterol-binding protein and localization of the gene to
 A:Reference number: A34581; MUID:90243258
 A:Accession: A34581
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-807 <LEV>
 A:Cross-references: GB:M86917; NID:g189402; PIDN:AAA59973.1; PID:g189403; GB:J04757
 C:Genetics:
 A:Gene: GDB:OSBP
 A:Cross-references: GDB:120252; OMIM:167040
 A:Map position: 11q12-11q13
 C:Superfamily: pleckstrin repeat homology

Query Match 15.8%; Score 74; DB 2; Length 807;
 Best Local Similarity 27.7%; Pred. No. 2.5;
 Matches 26; Conservative 13; Mismatches 27; Indels 28; Gaps 4;

QY 9 KGLQRMFLENDVLLY-----TSRGLTASNQFKVHGOLPLYGMTIESEDEWGV 59
 DB 103 KGYRRWFVLSNGLLSYYSKAEHRHCTCGTI-----NLATANIIVEDS----- 146
 QY 60 HC--LTLRGQRQSIIVAASSRSEMEKVEDIQMA 91
 DB 147 -CNFIISNGGAQTYHLKASSEVERQRWVTALELA 179

Search completed: December 6, 2001, 08:52:40
 Job time: 372 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:53:10 ; Search time 28.69 Seconds
(without alignments)
116.295 Million cell updates/sec

Title: US-09-555-342A-2_COPY_764_854

Perfect score: 468

Sequence: 1 GSLKSLGKGLQRMFFLEN.....IVAASSRSEKWKVEDIQMA 91

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109.5	23.4	266	1 YOTB_CAEEL	P34657 caenorhabdi
2	107	22.9	961	1 FGD1_HUMAN	P98174 homo sapien
3	106	22.6	960	1 FGD1_MOUSE	P52734 mus musculus
4	98	20.9	646	1 Y142_HUMAN	Q14155 homo sapien
5	84.5	18.1	394	1 CYH4_HUMAN	Q9u1a0 homo sapien
6	79.5	17.0	1244	1 GNRP_RAT	P28818 rattus norv
7	79	16.9	994	1 MYSA_DICDI	P22467 dictyosteli
8	76.5	16.3	1275	1 GNRP_HUMAN	Q13972 homo sapien
9	75	16.0	400	1 CYH2_HUMAN	Q99418 homo sapien
10	75	16.0	400	1 CYH2_MOUSE	P97695 mus musculus
11	74	15.8	807	1 OXYB_HUMAN	P22059 homo sapien
12	74	15.8	809	1 OXYB_RABIT	P16258 oryctolagus
13	72.5	15.5	1262	1 GNRP_MOUSE	P27671 mus musculus
14	71	15.2	398	1 CYH1_HUMAN	Q15438 homo sapien
15	71	15.2	753	1 YB2_YEAST	P38261 saccharomyc
16	70.5	15.1	399	1 CYH3_MOUSE	O08967 mus musculus
17	70	15.0	400	1 CYH3_HUMAN	O43739 homo sapien
18	70	15.0	400	1 CYH3_RAT	P97696 rattus norv
19	69.5	14.9	537	1 P4H2_MOUSE	Q60716 mus musculus
20	69	14.7	398	1 CYH1_MOUSE	Q9qx11 mus musculus
21	69	14.7	398	1 CYH1_RAT	P97694 rattus norv
22	67.5	14.4	689	1 ARK1_HUMAN	P25098 homo sapien
23	67	14.3	638	1 Y053_HUMAN	P42331 homo sapien
24	66.5	14.2	689	1 ARK1_BOVIN	P21146 bos taurus
25	66.5	14.2	689	1 ARK1_MESAU	Q64682 mesocricetu
26	65.5	14.0	409	1 Y148_MYCCE	P47394 mycoplasma
27	64	13.7	1332	1 SOS2_HUMAN	Q07890 homo sapien
28	63	13.5	1297	1 SOS2_MOUSE	Q02384 mus musculus
29	61	13.0	275	1 DAPF_CHLTR	O84437 chlamydia t
30	61	13.0	513	1 SYN_SYNY3	P52276 synecocyst
31	60.5	12.9	207	1 NEF_HV1ND	P18801 human immun
32	60.5	12.9	397	1 PEL_TOBAC	P40972 nicotiana t
33	60.5	12.9	860	1 ENV_HV2BE	P18094 human immun

RESULT 1

ID	YOTB_CAEEL	STANDARD;	PRT;	266 AA.
AC	P34657;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	HYPOTHETICAL 30.2 KDA PROTEIN ZK632.12 IN CHROMOSOME III.			
GN	ZK632.12.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE=94150718; PubMed=7906398;			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,			
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,			
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,			
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,			
RA	Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,			
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,			
RA	Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,			
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,			
RA	Waterson R., Watson A., Weinstein L., Wilkinson-Sproat J.,			
RA	Wohlman P.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans."			
RL	Nature 368:32-38(1994).			
CC	-!- SIMILARITY: CONTAINS 1 PH DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z22181; CA880187.1;			
DR	WormPep; ZK632.12; CE01110.			
DR	InterPro; IPR001849; PH.			
DR	InterPro; IPR000306; znf_FYVE.			
DR	Pfam; PF01363; FYVE; 1.			
DR	Pfam; PF00169; PH; 1.			
DR	SMART; SM00064; FYVE; 1.			
DR	SMART; SM00233; PH; 1.			
DR	PROSITE; PS50178; ZF_FYVE; 1.			
DR	PROSITE; PS50003; PH_DOMAIN; 1.			
KW	Hypothetical protein; Zinc-finger.			
FT	DOMAIN 35 131			
FT	ZN_FING 152 212 FYVE-TYPE.			
SQ	SEQUENCE 266 AA; 30187 MW; 91C2F62EDF13839E CRC64;			

P47807 gallus gall
P46735 mus musculus
Q05096 rattus norv
P16678 escherichia
Q06402 arabidopsis
Q07071 rattus norv
P04604 human immun
P32915 saccharomyc
P28817 rattus norv
P29851 streptococc
Q12774 homo sapien
P26819 rattus norv

ALIGNMENTS

```
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; Zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 590 689 PH 1.
FT ZN_FING 730 790 FYVE-TYPE.
FT DOMAIN 821 921 PH 2.
SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match      22.9%; Score 107; DB 1; Length 961;
Best Local Similarity 33.0%; Pred. No. 0.00018;
Matches 30; Conservative 17; Mismatches 40; Indels 4; Gaps 3;

QY 1 GSLSKLSGKG--LQORMFLEFNDVLLYTSRGLTASNQ-FKVHGQLPLYGMTIESEDEWG 57
DB 595 GHILKLSAKNGTTQDRYLILFNDRLLYCVPRLLRLLGKFSVRARIDVDGMELKESN-LN 653

QY 58 VPHCLTLRGQRQSIIVAASRSSEMEKWVEDI 88
DB 654 LPTFTLVSGKQRSLEQLQARTEEEKKDWQAI 684

RESULT 3
FGDL_MOUSE
ID FGDL_MOUSE STANDARD; PRT; 960 AA.
AC P52734;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).
GN FGDL_MOUSE
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96081343; PubMed=8535076;
RA Pastoris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,
RA Gorski J.L.;
RT "Cloning and regional localization of the mouse facio-genital
RT dysplasia (fgdl) gene.";
RL Mamm. Genome 6:658-661(1995).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GDP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U11690; AAA57004.1; -
CC MIM; 305400; -
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR000822; Znf-C2H2.
CC Pfam; PF01363; FYVE; 1.
CC Pfam; PF01169; PH; 2.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00064; FYVE; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00325; RHOGEF; 1.
CC PROSITE; PS50178; ZF_FYVE; 1.

Query Match      23.4%; Score 109.5; DB 1; Length 266;
Best Local Similarity 30.3%; Pred. No. 2e-05;
Matches 27; Conservative 21; Mismatches 40; Indels 1; Gaps 1;

QY 1 GSLSKLSGKGLOQRMFFLEFNDVLLYTSRGLTASNQFVKVHGQLPLYGMTIESEDEWVPH 60
DB 40 GVLVWCKRKPKQKQFFLEFNDILVY-GNIVISKRYNKQIRLKGQVQVEDLEDGIGKH 98

QY 61 CLTLRGQRQSIIVAASRSSEMEKWVEDIQ 89
DB 99 GWIITKPAKSFAYAAETETEKREWMLHIE 127

RESULT 2
FGDL_HUMAN
ID FGDL_HUMAN STANDARD; PRT; 961 AA.
AC P98174;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN).
GN FGDL_HUMAN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95042764; PubMed=7954831;
RA Pastoris N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,
RA Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.;
RT "Isolation and characterization of the facio-genital dysplasia
RT (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide
RT exchange factor.";
RL Cell 79:669-678(1994).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GDP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA,
CC AND LESS IN LIVER; ADULT HEART, BRAIN, LUNG, SKELETAL MUSCLE, AND
CC LESS IN PANCREAS AND LIVER.
CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE
CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT
CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U11690; AAA57004.1; -
CC MIM; 305400; -
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR000822; Znf-C2H2.
CC Pfam; PF01363; FYVE; 1.
CC Pfam; PF01169; PH; 2.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00064; FYVE; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00325; RHOGEF; 1.
CC PROSITE; PS50178; ZF_FYVE; 1.
```



```
DR PROSITE; PS50178; 2F_FYVE; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; Zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 589 688 PH 1.
FT ZN_FING 729 789 FYVE-TYPE.
FT DOMAIN 820 920 PH 2.
SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 22.6%; Score 106; DB 1; Length 960;
Best Local Similarity 33.0%; Pred. No. 0.00023;
Matches 30; Conservative 17; Mismatches 40; Indels 4; Gaps 3;

QY 1 GSLSKLSGKG--LQORMFLENDVLLYTSRGLTASNQ-FKVGQPLPLGYMTIESEDEWG 57
Db 594 GHILKUSANGTQDRYLILFNDRLLYCVPRLLGKFTVRARIDVDGMELKESN-LN 652

QY 58 VPHCLRLGRQSIIVAAASSRSEMEKRWEDI 88
Db 653 MPRTFLVSGKQSRLEQARTEEKKDWQAI 683

RESULT 4
Y142_HUMAN
ID Y142_HUMAN STANDARD; PRT; 646 AA.
AC Q14155;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0142.
GN KIAA0142.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res 2:167-174(1995).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63476; BAA09763.1; -
DR HSSP; P19174; LHSQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR00219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Hypothetical protein; SH3 domain.
FT DOMAIN 6 65
```

```
FT DOMAIN 295 400 PH.
SQ SEQUENCE 646 AA; 73140 MW; B5B5A83F0EBC28D2 CRC64;

Query Match 20.9%; Score 98; DB 1; Length 646;
Best Local Similarity 32.1%; Pred. No. 0.0012;
Matches 25; Conservative 17; Mismatches 34; Indels 2; Gaps 2;

QY 13 QRMFFLENDVLLYTSRGLTASNQFKVHGQPLPLGYMTIESEDEWGPVPHCLTLRQQR-QSI 71
Db 320 ERYLLLPNVLMLSLASPRMSG-FIYQGLPTTGMITKLEDSHNHNAFEISGSMIERI 378

QY 72 IVAASSRSEMEKRWEDIQ 89
Db 379 LVSCNNQDLQEWVHLQ 396

RESULT 5
CYH4_HUMAN
ID CYH4_HUMAN STANDARD; PRT; 394 AA.
AC Q9UIA0; Q9UGT6;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 4.
GN PSCD4 OR CYT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Liu D., Zhang H., Lu J.;
RT "cDNA cloning and genomic organization of cytohesin-4";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20119275; PubMed=10652308;
RA Ogasawara M., Kim S.C., Adamik R., Togawa A., Ferrans V.J., Takeda K.,
RA Kirby M., Moss J., Vaughan M.;
RT "Similarities in function and gene structure of cytohesin-4 and
RT cytohesin-1, guanine nucleotide-exchange proteins for ADP-ribosylation
RT factors.";
RL J. Biol. Chem. 275:3221-3230(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clamp M., Smink L.J., Alnsough R., Almeida J.P., Babbage A.,
RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Coile C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Fleming K French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA Matthews L., McCann O.T., Mcclay J., McLaren S., Mcmuray A.A.,
RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
```

RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Mala J. E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA Sloyan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA Minx P., Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H.,
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA Nelson J., Korn I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF1 AND ARF5.
CC PROMOTES THE ACTIVATION OF ARF THROUGH REPLACEMENT OF GDP WITH
CC GTP.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PERIPHERAL BLOOD
CC LEUKOCYTES.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF075458; AAF15389.1; -;
DR EMBL: AF123349; AAF28896.1; -;
DR EMBL: Z94160; AAG3067.1; -;
DR HSSP: Q99418; LPBV.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000904; Sec7.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF01369; Sec7; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00222; Sec7; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50190; SEC7; 1.
KW Guanine-nucleotide releasing factor; Coiled coil.
FT DOMAIN 12 65 COILED COIL (POTENTIAL).
FT DOMAIN 72 201 SEC7.
FT DOMAIN 259 375 PH.
SQ SEQUENCE 394 AA; 45672 MW; 0C9C8E9AD6890F34 CRC64;

Query Match 18.1%; Score 84.5; DB 1; Length 394;
Best Local Similarity 25.7%; Pred. No. 0.025;
Matches 29; Conservative 22; Mismatches 33; Indels 29; Gaps 5;

QY 1 GSLSKLSG--KGLQRMFFLNDVLLYTSRGLTASNQFKVHGOLPLYGMTIESEDEWG 58
Db 264 GLLKLGGRVKTWKRWFLITDNCLY-----FEFTDKPRGIIPLENLSVQKVDPP-KK 318
QY 59 PHCLTL-----RQQR-----QSIIVAASSRSEMEKWKVEDIQ 89
Db 319 PFCLELYNPSCRQKTKACKTDGDRVVEGKESYRISATSABERDQWIESIR 371

RESULT 6
GNRP_RAT STANDARD; PRT; 1244 AA.
AC P28818;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (P140 RAS-GRF).

GN RASGRF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92350260; PubMed=1379346;
RA Shou C., Farusworth C.L., Neel B.G., Feig L.A.;
RT "Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing
RL factor for Ras p21.";
RL Nature 358:351-354(1992).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67241; CAA47666.1; -;
DR PIR: S29083; S29083. GDS_CDC24.
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEF.
DR InterPro: IPR000219; RHOGEF.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF00169; PH; 2.
DR Pfam: PF00617; RasGEF; 1.
DR Pfam: PF00618; RasGEFN; 1.
DR Pfam: PF00621; RHOGEF; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00233; PH; 2.
DR SMART: SM00147; RasGEF; 1.
DR SMART: SM00229; RasGEFN; 2.
DR SMART: SM00325; RHOGEF; 1.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PROSITE: PS00741; GDS_CDC24; 1.
DR PROSITE: PS50003; PH_DOMAIN; 2.
DR PROSITE: PS50096; IQ; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 22 129 PH 1.
FT DOMAIN 204 229 IQ.
FT DOMAIN 244 455 DH.
FT DOMAIN 456 582 PH 2.
FT DOMAIN 1007 1241 RASGEF.
SQ SEQUENCE 1244 AA; 142666 MW; 4B647879E842AF6B CRC64;

Query Match 17.0%; Score 79.5; DB 1; Length 1244;
Best Local Similarity 26.9%; Pred. No. 0.37;
Matches 29; Conservative 14; Mismatches 38; Indels 27; Gaps 3;

QY 1 GSLSKLSGKGLQRMFFLNDVLLYTSRGLTASNQFKVHGOLPLYGMTI---ESEEDEWG 57
Db 477 GRGLSLKLKKEGRCFLFSKGLHLTKNGVISLIDCTLDLDDFNWDDG 536
QY 58 VPHCLTLRQQR-----SIIVAASSRSEMEKWKVEDI 88
Db 537 -----KQGEVDHDLFKIWVEPKDPPFTVILVASSRQKAAWTSDI 577

RESULT 7
MYSALDICI

```

ID MYSA_DICDI STANDARD; PRT; 994 AA.
AC P22467;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN IA HEAVY CHAIN (MYOSIN-LIKE PROTEIN ABMA).
GN ABMA OR DMTA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A5-2;
RX MEDLINE=92096486; PubMed=2519618;
RA Titus M.A., Warrick H.M., Spudich J.A.;
RT "Multiple actin-based motor genes in Dictyostelium.";
RL Cell Regul. 1:55-63(1989).
CC -!- FUNCTION: ACTIN-BASED MOTOR PROTEIN, POSSIBLY INVOLVED IN A WIDE
CC RANGE OF MOTILE PROCESSES, SUCH AS CELL MOVEMENT ACROSS A SURFACE,
CC AND EXTENSION AND RETRACTION OF PSEUDOPODIA OR LAMELLIPODIA.
CC -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S73909; AAB20711.1; -.
CC PIR; JQ0151; JQ0151.
CC HSP; P08799; 1MND.
CC DictyDb; DD01028; abma.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00612; IQ; 2.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; Actin-binding; ATP-binding; Calmodulin-binding;
CC Multigene family.
CC FT DOMAIN 1 743 MYOSIN HEAD-LIKE.
CC FT DOMAIN 744 773 IQ.
CC FT DOMAIN 774 994 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
CC FT NP_BIND 105 112 ATP (POTENTIAL).
CC FT DOMAIN 575 655 ACTIN-BINDING.
CC FT DOMAIN 830 931 MEMBRANE-BINDING (POTENTIAL).
CC SEQUENCE 994 AA; 113286 MW; D30A5D20885B118C CRC64;

Query Match 16.9%; Score 79; DB 1; Length 994;
Best Local Similarity 24.5%; Pred. No. 0.32;
Matches 25; Conservative 24; Mismatches 31; Indels 22; Gaps 5;

QY 3 LSKLSKGLQORMFNDVLLYTSRG--LTASNQKFKVHGQLPLYG---MTIESEDEWG 57
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 20 LTEVSESSLHENLKIRYKEGLIYTSIGPVLVSMNPK---QLGIYNDQINLYKGRHEFE 76

QY 58 VP-----HCLTLRGQRSIIIVASS-----RSEMEKVV 85
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 77 IPPIIYIADKAYRALRURSEGNOCIIISGSGAGKTEASKYI 118

RESULT 8
GNRP_HUMAN STANDARD; PRT; 1275 AA.
ID GNRP_HUMAN
```

```

AC Q13972;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE EXCHANGE FACTOR CDC25).
GN RASGRF1 OR CDC25.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei W., Broek D.;
RT "Cloning and analysis of the full length human cdc25 cDNA, a ras-
RT specific nucleotide exchange factor.";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L26584; AAA58417.1; -.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000651; RasGEFN.
CC InterPro; IPR001895; RasGRF_CDC25.
CC InterPro; IPR000219; RHOGEF.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF00169; PH; 2.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEEN; 1.
CC Pfam; PF00621; RHOGEF; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00147; RasGEF; 1.
CC SMART; SM00229; RasGEFN; 2.
CC SMART; SM00325; RHOGEF; 1.
CC PROSITE; PS00720; GDS_CDC25; 1.
CC PROSITE; PS00741; GDS_CDC24; 1.
CC PROSITE; PS50003; PH_DOMAIN; 2.
CC PROSITE; PS50096; IQ; 1.
CC Guanine-nucleotide releasing factor.
CC FT DOMAIN 22 129 PH 1.
CC FT DOMAIN 204 229 IQ.
CC FT DOMAIN 247 431 DH.
CC FT DOMAIN 467 584 PH 2.
CC FT DOMAIN 1038 1272 RASGEF.
CC SEQUENCE 1275 AA; 145381 MW; 86C6F54AA1E451F1 CRC64;

Query Match 16.3%; Score 76.5; DB 1; Length 1275;
Best Local Similarity 25.2%; Pred. No. 0.85;
Matches 28; Conservative 14; Mismatches 38; Indels 31; Gaps 3;

QY 1 GSKLSKGLQORMFNDVLLYTSRGLTASNQKFKVHGQLPLYGNTI-----ESEDEW 56
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 477 GRGLSLSLEKGERQRCFLFSKHLIICTRGSGKGLHLTNGVISLIDCTLLEPEESTEE- 535

QY 57 GVPHCUTLRLGQRO-----SIIVAASSRSEMEKVVEDI 88
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 536 -----AKGSGQDIDHDFKIGVEPKDSPPTFVLVASSRQKAAWTSDI 579
```

```
RESULT 9
CYH2_HUMAN STANDARD; PRT; 400 AA.
AC Q99418; Q92958;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 2 (ARF NUCLEOTIDE-BINDING SITE OPENER) (ARNO PROTEIN) (ARF
DE EXCHANGE FACTOR).
GN PSCD2 OR ARNO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Chardin P., Paris S., Antony B., Robineau S., Bernaud-Dufour S.,
RA Jackson C.L., Chabre M.;
RT "A human exchange factor for ARF contains Sec7 and pleckstrin-homology
RT domains.";
RL Nature 384:481-484(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=98079021; PubMed=9417041;
RA Frank S.F., Upender S.K., Hansen S.H., Casanova J.E.;
RT "ARNO is a guanine nucleotide exchange factor for ADP-ribosylation
RT factor 6.";
RL J. Biol. Chem. 273:23-27(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 51-252.
RX MEDLINE=98135767; PubMed=9476900;
RA Mossessova E., Gulbis J.M., Goldberg J.;
RT "Structure of the guanine nucleotide exchange factor Sec7 domain of
RT human ARNO and analysis of the interaction with ARF GTPase.";
RL Cell 92:415-423(1998).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 51-252.
RX MEDLINE=98169075; PubMed=9510236;
RA Cherifis J., Menetrey J., Mathieu M., le Bras G., Robineau S.,
RA Beraud-Dufour S., Antony B., Chardin P.;
RT "Structure of the Sec7 domain of the Arf exchange factor ARNO.";
RL Nature 392:101-105(1998).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF1, ARF3 AND
CC ARF6. PROMOTES THE ACTIVATION OF ARF THROUGH REPLACEMENT OF GDP
CC WITH GTP.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99753; CAAG8084.1; -
DR EMBL; U70728; AAB09591.1; -
DR PDB; 1PBV; 09-WAR-99.
DR MIM; 602488; -
DR InterPro; IPR001849; PH.
DR InterPro; IPR000904; Sec7.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF01369; Sec7; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00222; Sec7; 1.
```

```
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50190; SEC7; 1.
KW Guanine-nucleotide releasing factor; Coiled coil;
KW Alternative splicing; 3D-structure.
FT DOMAIN 10 63 COILED COIL (POTENTIAL).
FT DOMAIN 72 201 SEC7.
FT DOMAIN 259 376 PH.
FT VARSPLIC 272 272 MISSING (IN ISOFORM 2).
SQ SEQUENCE 400 AA; 46546 MW; 70441A58483BDOE1 CRC64;

Query Match 16.0%; Score 75; DB 1; Length 400;
Best Local Similarity 24.3%; Pred. No. 0.32;
Matches 28; Conservative 21; Mismatches 38; Indels 28; Gaps 5;

QY 1 GSLSKLSG---KGLQRMFFLFNDVLLYTSRGLTASNQFKVHQQLPLYGWTIEESD--- 54
DB 264 GWLLKGGGRVKTWRKRWFFILTONCLYY-----FEYTTDKPRGIIPLENISIREVDDPRK 319
QY 55 ----EWGVPH-----CLT-----LRGQRQSIIVAAASRSSEMEKWKVEDIQMA 91
DB 320 PNCFELYIPNNKQLIKACKTEADGRVVEGNHMYRISAPTQEKDEWIKSIQAA 374

RESULT 10
CYH2_MOUSE STANDARD; PRT; 400 AA.
AC P97695; O89099;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 2 (ARF NUCLEOTIDE-BINDING SITE OPENER) (ARNO PROTEIN) (CLM2)
DE (SEC7 HOMOLOG B) (MSEC7-2).
GN PSCD2 OR SEC7B.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RC SPECIES=Mouse; TISSUE=Brain;
RX MEDLINE=98416066; PubMed=9744817;
RA Kim H.-S., Chen Y., Lonai P.;
RT "Complex regulation of multiple cytohesin-like genes in murine tissues
RT and cells.";
RL FEBS Lett. 433:312-316(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC SPECIES=Mouse;
RA Liu D., Zhang H., Lu J.;
RT "cDNA cloning of mouse cytohesin-2 and demonstration of its
RT association with the integrin beta2 subunit.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC SPECIES=Rat;
RX MEDLINE=98013473; PubMed=9352219;
RA Telmenakis I., Benseler F., Stenius K., Sudhof T.C., Brose N.;
RT Rat homologues of yeast sec7p.;
RL Eur. J. Cell Biol. 74:143-149(1997).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF1, ARF3 AND
CC ARF6. PROMOTES THE ACTIVATION OF ARF THROUGH REPLACEMENT OF GDP
CC WITH GTP.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1/CLM2-A (SHOWN HERE), 2 AND
CC 3/CLM2-B; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES TESTED, WITH HIGHEST
CC PROTEIN LEVELS IN BRAIN AND ADRENAL.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```



```
CC ----- J05056; AAA31427.1; -.
DR EMBL; A34404;
DR InterPro; IPR000648; Oxysterol_BP.
DR InterPro; IPR001849; PH.
DR Pfam; PF01237; Oxysterol_BP; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS01013; OSBP; 1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 92 GLY/ALA-RICH.
FT DOMAIN 90 183 PH.
SQ SEQUENCE 809 AA; 89478 MW; 55C6CAE1B985B1E0 CRC64;

Query Match 15.8%; Score 74; DB 1; Length 809;
Best Local Similarity 27.7%; Pred. No. 0.96;
Matches 26; Conservative 13; Mismatches 27; Indels 28; Gaps 4;

QY 9 KGLQRMFFELNDVLLY-----TSRGLTASNQFKVHGQLPLYGTMTESEDEWGV 59
DB 105 KGQRRWFLVSLNGLLSYRSKAEHRFTCGTI-----NLATANITVEDS----- 148

QY 60 HC--LTLRGQRQSIIVAASSRSEMEKVVEDIQMA 91
DB 149 -CNFIISNGGAQIVHLKASSEVERQRWVTALELA 181

RESULT 13
ID GNRP_MOUSE STANDARD; PRT; 1262 AA.
AC P27671;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE EXCHANGE FACTOR CDC25) (CDC25MM).
GN RASGRF1 OR CDC25 OR GRF1.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=BALB/C.
RX MEDLINE=93010996; PubMed=1396590;
RA Cen H., Lowy D.D.;
RT "Isolation of multiple mouse cDNAs with coding homology to
RT Saccharomyces cerevisiae CDC25: identification of a region related to
RT Bcr, Vav, Dbl and CDC34."
RL EMBO J. 11:4007-4015(1992).
RN [2]
RP SEQUENCE OF 791-1262 FROM N.A.
RC STRAIN=SWISS; TISSUE=Brain;
RX MEDLINE=92289680; PubMed=1376246;
RA Martegani E., Vanoni M., Zippel R., Coccetti P., Brambilla R.,
RA Ferrari C., Sturani E.P., Alberghina L.;
RT "Cloning by functional complementation of a mouse cDNA encoding a
RT homologue of CDC25, a Saccharomyces cerevisiae RAS activator."
RL EMBO J. 11:2151-2157(1992).
RN [3]
RP SEQUENCE OF 1031-1226 FROM N.A.
RX MEDLINE=92357779; PubMed=1379731;
RA Wei W., Mosceller R.D., Sanyal P., Gonzales E., McKinney D.,
RA Dasgupta C., Li P., Liu B.X., Broek D.;
RT "Identification of a mammalian gene structurally and functionally
RT related to the CDC25 gene of Saccharomyces cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
```

```
CC -----
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L20899; AAA02741.1; -.
DR EMBL; X59868; CAA42525.1; -.
DR PIR; S20730; S20730.
DR PIR; S22693; S22693.
DR MGD; MGI:99694; Rasgrf1.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGEF_CDC25.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS50096; IQ; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 22 130 IQ.
FT DOMAIN 208 233 IQ.
FT DOMAIN 248 459 DH.
FT DOMAIN 460 588 PH 2.
FT DOMAIN 1025 1259 RASGEF.
FT CONFLICT 1033 1033 E -> D (IN REF. 3).
SQ SEQUENCE 1262 AA; 144101 MW; 38BFE68F7C228DC8 CRC64;

Query Match 15.5%; Score 72.5; DB 1; Length 1262;
Best Local Similarity 27.2%; Pred. No. 2.4;
Matches 28; Conservative 13; Mismatches 47; Indels 15; Gaps 3;

QY 1 GSLSKSLGKGLQRMFFELNDVLLYTSRGLTASNQFKVHGQLPLYGWTI---EESSEW 56
DB 481 GRGLSLTKKEGRQCFLFSKHLIICITRGSGGKLHLTKNGVISLIDCTLDEPENLDEA 540

QY 57 -----GVPHCLTLRGQRQ-----SIIVAASSRSEMEKVVEDI 88
DB 541 KGAGPETEHLFEKIGVPEKDSLPTFTVLVASTRQEKAAWTSDI 583

RESULT 14
ID CYH1_HUMAN STANDARD; PRT; 398 AA.
AC Q15438; O9P123; O9P124;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 1 (SEC7 HOMOLOG B2-1).
GN PCSO1 OR D17S811E.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:54:35 ; Search time 77.11 seconds
(without alignments)
172.621 Million cell updates/sec

Title: US-09-555-342A-2_COPY_764_854

Perfect score: 468

Sequence: 1 GSIKSLGKGLQRMFLFN.....IVAASSRSEKWKVEDIQMA 91

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.17.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organalle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-virus.*
13: sp-vertebrate.*
14: sp-unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	468	100.0	Q94f1 homo sapien
2	461	98.5	Q99k03 mus musculus
3	287	61.3	Q94887 homo sapien
4	152.5	32.6	Q9nr8 homo sapien
5	152.5	32.6	Q9p215 homo sapien
6	149.5	31.9	Q9d3y7 mus musculus
7	122	26.1	O88387 rattus norv
8	119	25.4	Q93504 brachydanio
9	118.5	25.3	Q9ub9 homo sapien
10	116.5	24.9	Q9nm16 mus musculus
11	112	23.9	O88841 mus musculus
12	111.5	23.8	Q9h8w4 homo sapien
13	111.5	23.8	Q9vn69 drosophila
14	107	22.9	Q99k79 mus musculus
15	104	22.2	Q9neh0 leishmania
16	102	21.8	O76902 drosophila
17	99.5	21.3	O18284 caenorhabdi
18	99	21.2	Q9nxy1 homo sapien
19	96	20.5	O55043 rattus norv

20	94.5	20.2	414	5	Q20653	Q20653 caenorhabdi
21	88	18.8	1028	4	Q9UGQ4	Q9ugq4 homo sapien
22	88	18.8	1539	4	Q9P2D2	Q9p2d2 homo sapien
23	87	18.6	630	11	Q9ES27	Q9es27 mus musculus
24	87	18.6	646	11	O08757	O08757 mus musculus
25	87	18.6	705	11	Q9ES28	Q9es28 mus musculus
26	86	18.4	733	11	O88842	O88842 mus musculus
27	85.5	18.3	762	4	Q9H961	Q9h961 homo sapien
28	84.5	18.1	740	4	Q9UPP2	Q9upp2 homo sapien
29	81.5	17.4	201	4	Q9NUL6	Q9nul6 homo sapien
30	81	17.3	211	4	Q9UFW5	Q9ufw5 homo sapien
31	81	17.3	301	4	Q9H9R2	Q9h9r2 homo sapien
32	81	17.3	408	4	Q9H9D7	Q9h9d7 homo sapien
33	80.5	17.2	554	11	Q9JHT9	Q9jht9 mus musculus
34	80	17.1	755	11	Q9CXJ3	Q9cxj3 mus musculus
35	79	16.9	1355	4	Q9HCK9	Q9hck9 homo sapien
36	78.5	16.8	875	5	Q18372	Q18372 caenorhabdi
37	78	16.7	400	11	Q9EQD2	Q9equ2 apodemus ag
38	77	16.5	496	4	Q9NTG0	Q9ntg0 homo sapien
39	77	16.5	619	4	Q9HDC6	Q9hdc6 homo sapien
40	77	16.5	694	4	Q9UPP0	Q9upp0 homo sapien
41	77	16.5	720	4	Q9NR80	Q9nr80 homo sapien
42	77	16.5	1210	5	Q9VP80	Q9vp80 drosophila
43	75.5	16.1	399	11	Q99KH2	Q99kh2 mus musculus
44	75.5	16.1	1194	4	Q9P2F6	Q9p2f6 homo sapien
45	75	16.0	266	4	Q15795	Q15795 homo sapien

ALIGNMENTS

RESULT 1

Q9Y4F1 PRELIMINARY; PRT; 1045 AA.
AC Q9Y4F1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE CDEP;
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RX MEDLINE=98086358; PubMed=9425278;
RA Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,
RA Kato Y.;
RT "Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors.";
RL Biochem. Biophys. Res. Commun. 241:369-375(1997).
DR EMBL; AB008430; BAA24267.1; -;
DR HSSP; P08567; IPLS.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00621; RhoGEF; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00560; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS00557; BAND_41_3; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
SQ SEQUENCE 1045 AA; 118632 MW; 0E8B2D61C0F58417 CRC64;

Query Match 100.0%; Score 468; DB 4; Length 1045;
Best Local Similarity 100.0%; Pred. NO. 1.5e-46;

QY 1 GSKLSGKGLQQRMFLEFNDVLLYTSRGLTASNQFKVHGQLPLYGMTIEESEDEWGVPH. 60

RESULT 7

Y

```

088387
ID O88387 PRELIMINARY; PRT; 766 AA.
AC O88387;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ACTIN-FILAMENT BINDING PROTEIN FRABIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98334590; PubMed=96680039;
RA Obaishi H., Nakanishi H., Mandai K., Satoh K., Satoh A., Takahashi K.,
RA Miyahara M., Nishiooka H., Takaiishi K., Takai Y.,
RT "Frabin, a novel FGD1-related actin filament-binding protein capable
RT of changing cell shape and activating c-Jun N-terminal Kinase.";
RL J. Biol. Chem. 273:18697-18700(1998).
DR EMBL; AF038388; AAC27698.1; -.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR000306; Znf_FYVE.
DR InterPro; IPR000822; Znf-C2H2.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF01363; FYVE; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhoGEF; 1.
SQ SEQUENCE 766 AA; 86453 MW; E3481DC0B0B334C CRC64;

Query Match 26.1%; Score 122; DB 11; Length 766;
Best Local Similarity 31.5%; Pred. No. 7.4e-06;
Matches 29; Conservative 24; Mismatches 35; Indels 4; Gaps 3;

QY 1 GSLKSLGCK--GLQRMFFLENDVLLY--TSRGLTASNQFKVHGOLPLYGMYTIESEDEWG 57
Db |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
427 GQILKLAAINTSAQERYLLFNNMLLYCVPRFSLVSGFTVTRVGDGMKIVETHNE-E 485

QY 58 VPHCLITLRGQRQSIIVAAASSRSEMKWVEDIQ 89
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 486 YPHFQVSGKERTLELQASSQDKKEWIKALQ 517

RESULT 8
O93504 PRELIMINARY; PRT; 621 AA.
AC O93504;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FACIOGENITAL DYSPLASIA PROTEIN.
GN FGD.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX Pasteris N.G., Buckler J., Gorski J.L.;
RT "Isolation and Characterization of a zebrafish homologue of the
RT FacioGenital Dysplasia (FGD1) Gene.";
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017370; AAC35432.1; -.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR000306; Znf_FYVE.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.

```

[illegible]

```

SQ SEQUENCE 279 AA; 31200 MW; 14F3B2B9AAF96DD5 CRC64;

Query Match
Best Local Similarity 24.9%; Score 116.5; DB 11; Length 279;
Matches 27; Conservative 28; Mismatches 27; Indels 13; Gaps 3;

QY 1 GSLSKLSGKGLQORWFFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTE-----ESED 54
DB 40 GVLTRECKRKAPRIFFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTE-----ESED 54
QY 55 ENGVPCHLTQRQOSIIVAASSRSEMEKWKVEDIQ 89
DB 99 RW-----MIKTAKKSFVVSAASTTERQEWISHIE 127

RESULT 11
O88841 PRELIMINARY; PRT; 727 AA.
AC O88841
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE FACIOGENITAL DYSPLASIA PROTEIN 2.
GN FGD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Pasteris N.G., Gorski J.L.;
RT "Isolation, characterization and mapping of Fgd2, a homologue of the
RT facio-genital dysplasia (FGD1; Aarskog syndrome) gene.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF017368; AAC35430.1; -.
DR MGD: MGI:1347084; Fgd2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhoGEF; 1.
DR PROSITE: PS00003; PH DOMAIN; 1.
SQ SEQUENCE 727 AA; 82099 MW; 5CD0FBFB905C5FC5 CRC64;

Query Match
Best Local Similarity 23.9%; Score 112; DB 11; Length 727;
Matches 27; Conservative 17; Mismatches 34; Indels 2; Gaps 2;

QY 13 QRMFFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTEESDEGVPHCLTLRQQRQSI 71
DB 338 ERYLVFNMLLYCVRVLQVGAQVQVTRIDVAGMKVRELTDVAFPHSFLVSGKQRTL 396
QY 72 IVAASSRSEMEKWKVEDIQ 91
DB 397 ELQARSDEMSWMAQCA 416

RESULT 12
O9H8W4 PRELIMINARY; PRT; 249 AA.
AC O9H8W4
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CDNA FLJ13187 FIS, CLONE NT2RP3004242, WEAKLY SIMILAR TO PUTATIVE
DE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK023249; BAB14486.1; -.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS00003; PH DOMAIN; 1.
SQ SEQUENCE 249 AA; 27797 MW; F5E3F84595A98886 CRC64;

Query Match
Best Local Similarity 23.8%; Score 111.5; DB 4; Length 249;
Matches 28; Conservative 20; Mismatches 39; Indels 1; Gaps 1;

QY 1 GSLSKLSGKGLQORWFFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTEESDEGVPH 60
DB 40 GVLTCLCRKKRPFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTEESDEGVPH 98
QY 61 CLTLRQQRQSIIVAAASSRSEMEKWKVEDI 88
DB 99 GWLIKTPTKSFVYAATATEKSEMMNHI 126

RESULT 13
Q9VN69 PRELIMINARY; PRT; 562 AA.
AC Q9VN69
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CG2008 PROTEIN.
GN CG2008.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

[illegible]

QY	1	GSLSKLSGKG--LQRMFFLFNFDVLLTYTSRGLTASNQ-FKVHGQLPLYGMTIEESEDWNG	57
.	:	: : :	:
Db	31	GHTLKSAKNGTQDHYLLFNDRLLVCVPRLRGKFVSRAIDVDGMELKESSN-LN	89
.	:	: : :	:
QY	58	VPHCTLRGQRQSIIVAASSRSEMEKKWEDI	88
.	:	: : :	:

Query Match	22.2%	Score 104;	DB 5;	Length 1161;
Best Local Similarity	29.0%	Pred. NO.	0.0016;	
Matches	27:	Conservative	23:	Mismatches 37:
				Indels 6:
				Gaps 3:

Search completed: December 6, 2001, 08:54:36
Job time: 278 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:51:42 ; Search time 72.4 Seconds
(without alignments)
198.484 Million cell updates/sec

Title: US-09-555-342A-2_COPY_544_737

Perfect score: 1023

Sequence: 1 IAEVSTERTYKDLKLEVIT.....AEITEMVAQLHGTMRKMFN 194

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	100.0	1045	20 AAY07482	Human chondrocyte-
2	1023	100.0	1045	21 AAY91947	Human cytoskeleton
3	924	90.3	395	21 AAB54227	Human pancreatic c
4	195.5	19.1	766	21 AAY51248	Rat actin-binding
5	176	17.2	523	22 AAM39338	Human polypeptide
6	176	17.2	647	22 AAM41124	Human polypeptide
7	171	16.7	619	22 AAB97025	Human colon carcin
8	169	16.5	1683	21 AAY71160	Rat phosphodiester
9	168.5	16.5	1658	21 AAY57450	Mouse Ese2L protei
10	163	15.9	1715	21 AAY57449	Mouse Esi1L protei
11	160.5	15.7	844	21 AAB36516	Candida albicans C

12	149	14.6	1520	20 AAY41010	Amino acid sequenc
13	147	14.4	1527	22 AAU01184	Rat glutamate tran
14	138	13.5	2861	18 AAW27227	Human TRIO phospho
15	131	12.8	1461	19 AAW64468	Human secreted pro
16	131	12.8	1461	22 AAB90743	Human CW420.2 pro
17	125	12.2	846	22 AAM41561	Human polypeptide
18	122.5	12.0	1237	22 AAM39825	Human polypeptide
19	122.5	12.0	1257	22 AAM41611	Human polypeptide
20	122	11.9	580	20 AAW81349	Human guanine nucl
21	120	11.7	1244	21 AAY68825	Amino acid sequenc
22	118.5	11.6	798	22 AAB94519	Human protein seqd
23	118	11.5	477	15 AAB66067	Phosphorylated p66
24	117.5	11.5	1227	22 AAB65637	Novel protein kina
25	117.5	11.5	1289	20 AAY27163	Peptide Seq ID No:
26	117.5	11.5	1289	21 AAY56781	Human Trad protein
27	113.5	11.1	1269	22 AAM40297	Human polypeptide
28	113.5	11.1	1315	22 AAM42083	Human polypeptide
29	110	10.8	550	20 AAW81351	Human guanine nucl
30	110	10.8	1604	22 AAB47327	FCR4. Homo sapie
31	108	10.6	2596	22 AAB30569	A splice variant o
32	106	10.4	845	21 AAY49419	PKA substrate, Vav
33	105	10.3	827	22 AAM40424	Human polypeptide
34	105	10.3	827	22 AAM40425	Human polypeptide
35	105	10.3	872	20 AAY07073	Renal cancer assoc
36	104	10.2	391	21 AAB37413	Human secreted pro
37	103.5	10.1	847	20 AAY22237	Human KDR signal t
38	102	10.0	408	22 AAB95198	Human protein sequ
39	98	9.6	378	21 AAB54156	Human pancreatic c
40	98	9.6	596	21 AAB01203	Human GTPase assoc
41	97	9.5	235	20 AAW81350	Human guanine nucl
42	97	9.5	327	20 AAW73380	CSB2 protein sequ
43	97	9.5	384	21 AAB37414	Human secreted pro
44	97	9.5	412	22 AAB95879	Human protein sequ
45	97	9.5	526	20 AAW73378	CSB2 protein sequ

ALIGNMENTS

RESULT 1
AAV07482
ID AAY07482 standard; Protein; 1045 AA.
XX AAY07482;
AC AAY07482;
DT 17-AUG-1999 (first entry)
XX Human chondrocyte-derived protein CDEP.
DE
XX
KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW Db1 homology domain; plectstrin homology domain; rheumatoid 'arthritis;
KW drug.
XX
OS Homo sapiens.
XX
PN WO9928458-Al.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-JP05348.
XX
PR 27-NOV-1997; 97JP-0342060.
XX (CHUS) CHUGAI SEIYAKU KK.
XX
PI Kato Y, Kawamoto T, Koyano Y;
XX
DR WPI; 1999-371117/31.
XX N-PSDB; AAX79183.
XX
PT Protein CDEP expressed in differentiated chondrocytes, and gene
PT encoding it
XX

[illegible]

Modified-site	966	Human pancreatic cancer antigen protein sequence SEQ ID NO:679.
FT	WT	Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; cytostatic; neuroprotective; neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological; antiinflammatory; cardiant; gene therapy; chromosome mapping; linkage analysis; tissue identification; tissue typing; forensic; neural; immune system; muscular; reproductive; gastrointestinal; pulmonary; cardiovascular; renal; proliferative.
XX	XX	Homo sapiens.
XX	XX	WO200055320-A1.
XX	XX	21-SEP-2000.
XX	XX	08-MAR-2000; 2000WO-US05989.
XX	XX	12-MAR-1999; 99US-0124270.
XX	XX	(HUMA-) HUMAN GENOME SCI INC.
XX	XX	Rosen CA, Ruben SM;
XX	XX	WPI; 2000-579444/54.
XX	XX	N-PSDB; AAC98992.
XX	XX	New nucleic acid that is a pancreatic cancer antigen for preventing, treating, or ameliorating a medical condition, particular pancreatic cancer, or for use in assays for diagnosing a pathological condition - Claim 11; Page 1115-1116; 1379pp; English.
XX	XX	AAC98773 to AAC99231 encode the human pancreatic cancer associated proteins, called pancreatic cancer antigens, given in AAB54008 to AAB54466. The human pancreatic cancer antigens have cytostatic, neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive, gynaecological, cardiant and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer. Agonists and antagonists to the antigens can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of antibodies and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent sequences used in the exemplification of the present invention.
XX	XX	Sequence 395 AA;
XX	XX	Query Match 90.3%; Score 924; DB 21; Length 395;
XX	XX	Best Local Similarity 100.0%; Pred. No. 9.6e-89;
XX	XX	Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	XX	22 WFOSTVSKEDAMPEALKSLIFPNFPEPLHFKHTNFKLEQLALWEGRSNAQIRDYORIG 81
XX	XX	21 wfgstvsksedampealkslifpnfepihkftnlklegqlalwegrnaqirdygrig 80
XX	XX	82 DVMLKNIQGMKHLAAHLWKHSEALEALENGKSSRRLENFCRDFELQKVCYPLPNTFLR 141
XX	XX	81 dvmlknlqgmkhlaahlwkhsalealealengkssrrlenfcrdfelqkvcyplpntflr 140
XX	XX	142 PLHRLMHYKQVLERLCKHPPSHADFCRDLAALAEITEMVAQLHGTMKWFEN 194


```

Qy 120 NF--CRDFELQKVCYCLPLNTFLERLRLMHYKQVLERLCKHHPPSHADFRCRAALAEI 177
Db 227 ffeacr--llqgmldaidgflitpvgkickyplqlaeellkytaqhsdryyvaalavm 284

Qy 178 TEMVAQLHGHTMKRMEN 193
Db 285 rnvtaqinertkrrlen 300

```

RESULT 6
AAM41124
ID AAM41124 standard; Protein: 647 AA.

DT 22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 6055.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; ankytrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

OS : Homo sapiens.

XX PN WO200153312-A1.

26-JUL-2001.

26-DEC-2000;

PR 21-JAN-2000; 2000US-0488725.

PR 23-APR-2000; 2000US-0552317
PR 09-JUL-2000; 2000US-0598042

PR I9=JUL=2000; 2000US=0620312

PR 03=AUG=2000; 2000US=0653450

PR 14-SEP-2000; 2000US-0662191
PR 19-OCT-2000; 2000US-0693036

PR 29-NOV-2000; 2000US-0727344
yy

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

WPI: 2001-442253/47.

DR N-PSDB; AA160280.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 6055; 10078pp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with neurotrophic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

XX
SQ Sequence 647 AA;

Query Match 17.2%; Score 176; DB 22; Length 647;
Best Local Similarity 25.0%; Pred. No. 7.5e-10;
Matches 49; Conservative 43; Mismatches 92; Indels 1

Qy	1	IAKEVSTERTYKDL-EVITSNFQSVKSEDAMPEALKSLIPNPEPLHKFTNPLKEI	59
Db	238	vineimstehykhkldicegylkqcrkrdmsfdqglvfniedlyrfqmgfvrdl	297
Qy	60	EORLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALENGTKSSRRLE	119
Db	298	ekqy-----nnddphlseigpcflhqdgfwlyseycnnhldacmelsklmkdsryqh	350
Qy	120	NF--CRDFELQKVCYLPNTFLRLPLRLMHYKQVLERLCKHHPSHADPDCRAALAEI	177
Db	351	ffeasr--llqgmldaidgfllltpvqkikcypqlaelklyktaqhdshdyryvaalavm	408
Qy	178	TENVAQLHGTMIKMEN	193
Db	409	rnvtgaqnerkrillen	424

RESULT 7

AAB97025

ID AAB97025 standard; protein; 619 AA.

AA
AC AAB97025:XX
DT 16-JUL-2001 (first entry)

XX Human colon carcinoma suppressor gene-related protein.

XX Human; colon carcinoma suppressor gene; cytostatic;
KW
KW armadillo repeat; colon cancer.

XX Homo sapiens.

XX
PN
.TP2001057888-AXX
06-MAR-2001

XX 20-AUG-1999. 99TP-0234809

XX
20-AUG-1999. 99.TP-0234809

XX
PA (DAUC \ DATTCH PHARM CO LTDXX
DR WPT: 2001-310665/33

DR N-PSDB; AAF99965.

PT Colon carcinoma suppressor gene-related protein -

PS Claim 1: Fig 1: 20pp: Japanese.

ax The present sequence is a novel colon carcinoma suppressor gene-related
cc protein. The invention relates to the present sequence having a 619
cc residue amino acid sequence, to a sequence containing the present
cc sequence, and to a sequence being at least 70% homologous to the
cc present sequence. The protein binds to the armadillo repeat site
cc of the protein encoded by the cancer inhibiting gene of colon cancer.
cc It may be used in the treatment of colon tumours.

AA	Sequence	619	AA:
----	----------	-----	-----

Query Match 16.7%; Score 171; DB 22; Length 619;
Best Local Similarity 26.0%; Pred NC 2.4e-09;

QY 1 TAKEVSTTERTYLKDL-EVITSWFOSTVSKEDAMPEALKSLIFPNFEPLHKFHTNFLKEI 59

Db 217 vinsllsterdykhlrdicegyvrgkradmfsseqqlrtifgniediyrcqkafvkal 276
QY 60 EORLALWEGSRNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 119
Db 277 eqrf-----nrphlseigacflqhgdqfgyseycnnhpnacvelsrllklskyv 329
QY 120 NF--CRDFELQKVCYLPNTFLRLHRLMHYKQVLERLCKHHPHSHADFRDCRAALAEI 177
Db 330 ffeacr--llqkmidisldgflitpvcikcykplqlaelkkythpqrhdfkveaalham 387
QY 178 TEMVAQLHGTMIKMEN 193
Db 388 knvaqlinerkrrlen 403
RESULT 8
AAV71160
ID AAY71160 standard; Protein; 1683 AA.
XX AC AAY71160;
XX
DT 08-SEP-2000 (first entry)
XX DE Rat phosphodiesterase interacting protein, M14.
XX Rat; phosphodiesterase interacting protein; M14; PDE; CAMP-PDE;
KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
KW antiinflammatory; antipsoriatic; dermatological; antibacterial; shock;
KW analgesic; immunosuppressive; antidiabetic; vasotropic; antiarthritic;
KW antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;
KW antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;
KW atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
KW eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
KW reperfusion injury; atopic dermatitis; diabetes insipidus;
KW conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
KW arterial restenosis; ankylosing spondylitis; transplant rejection;
KW graft versus host disease.
XX
OS Rattus sp.
XX
XX WC200027861-A1.
XX
XX 18-MAY-2000.
XX
XX 12-NOV-1999; 99WO-US26860.
XX
XX 12-NOV-1998; 98US-0108255.
XX
XX (STRD) UNIV LELAND STANFORD JUNIOR.
XX
XX Conti M, Pahlke G;
XX
XX WPI; 2000-376479/32.
XX
XX Polynucleotide encoding a phosphodiesterase (PDE) interacting
PT polypeptide, useful for diagnosis and treatment of asthma, cystic
PT fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
XX Disclosure; Fig 6; 77pp; English.
XX
XX The present sequence is a phosphodiesterase (PDE) interacting protein,
CC M14 from rat. The protein modulates the functions and properties of PDEs,
CC specifically CAMP-PDEs, and also targets them to specific subcellular
CC compartments. The present sequence
CC can be used in the diagnosis and treatment of disease conditions
CC associated with PDE activity. The diseases include asthma, cystic
CC fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
CC granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
CC septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
CC inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
CC distress syndrome, diabetes insipidus, allergic rhinitis, allergic
CC conjunctivitis, vernal conjunctivitis, arterial restenosis,

CC atherosclerosis, inflammatory diseases associated with irritation and
CC pain, rheumatoid arthritis, ankylosing spondylitis, transplant
CC rejection and graft versus host disease, disease conditions associated
CC with hypersecretion of gastric acid, and disease conditions in which
CC cytokines are mediators.
XX
SQ Sequence 1683 AA;
Query Match 16.5%; Score 169; DB 21; Length 1683;
Best Local Similarity 24.2%; Pred. No. 1.5e-08;
Matches 48; Conservative 50; Mismatches 80; Indels 20; Gaps 5;
QY 4 EVSTTERTYKLEVITSWFQSTVSKEDAMPEALKSLIFPNFPLKHFNFLEKTEORL 63
Db 1202 elieteerymddqlvievfqkrmaesgfiteaemalifvnmkelmsntklkairv- 1260
QY 64 ALWEGSRNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCR 123
Db 1261 ----kktggekmpvemmgdilaee---lshmqyirfcscql----ngaallqktdeda 1309
QY 124 DFE--LQKVCY-----LPLNTELLRLHRLMHYKQVLERLCKHHPHSHADFRDCRAALA 175
Db 1310 drkeflkklasoprcgkmpslssfilkpmqritrypllrilrsilentpqnhvdhsslkiale 1369
QY 176 EITEMVAQLHGTMIKMEN 193
Db 1370 raeelcsqynegvren 1387
RESULT 9
AAV7450
ID AAY57450 standard; Protein; 1658 AA.
XX AC AAY57450;
XX
DT 28-FEB-2000 (first entry)
XX
XX Mouse Ese2L protein sequence.
XX
XX Mouse; murine; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
OS Mus sp.
XX
XX WO9955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX
XX 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
XX
XX WPI; 2000-052802/04.
XX
XX N-PSDB; AAZ39026, AAZ39027.
XX
XX New nucleic acid encoding Esel and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
XX Claim 33; Page 69-70; 99pp; English.
XX
XX The present invention specifically describes mammalian Esel and 2
CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-
CC mediated endocytosis (as a complex with Esp15 protein), vesicular

CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists, mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab); sequences, antisense to the (I) polynucleotide; agents that downregulate expression of ESE genes or antagonists of an ESE binding partner are used to treat diseases associated with undesirable endocytosis and resulting changes in cellular function. Particularly overexpression of ESE is used to block clathrin-mediated endocytosis in vivo or in cell cultures, while administration of (I) is used to promote endocytosis of selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of cells that can be stimulated to proliferate by a growth factor receptor; and similar compounds (also inactive ESE mutants) can be used to prevent viral infection. Endocytosis may also be regulated, in vivo or in cell cultures, by forming an ESE-Esp15 complex, then binding dynamin to the complex. Generally conditions that can be treated include cancer; abnormal cell division or migration; viral infection; or abnormal receptor signalling, tissue development or synaptic transmission. The present sequence represents mouse ESE21 protein sequence.

XX Sequence 1658 AA;

Query Match 16.5%; Score 168.5; DB 21; Length 1658;
Best Local Similarity 25.6%; Pred. No. 1.6e-08;
Matches 51; Conservative 48; Mismatches 79; Indels 21; Gaps 6;

QY 4 EVSTTERTYL-KDLEVITSWFQSTVSKEDAMPEALKSLIFPNFPELKHFTNFKIEQR 62

DB 1176 eliqteerymddqlqviefvqkrmaeegfitedmalifvfwkelmntklralrvr 1235

QY 63 LALWEGSRNAQIRDYQIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFC 122

DB 1236 -----kktggekmpvgnigdiiaae---ishmqayirfscql-----ngatllqqktdd 1283

QY 123 RDPE--LQKVY-----LPLNTFLRLPLRLHMLHYKQVLERLCKHPPSHADFRDRAAL 174

DB 1284 tdfkfkklasprckgplssflkpmqtrityrpllrslilentpqshvdhssiklal 1343

QY 175 AEITEMVAQLHGTWIKMEN 193

DB 1344 eraeicsgvnegvreken 1362

RESULT 10

AA57449

ID AA57449 standard; Protein; 1715 AA.

XX AAY57449;

XX AAY57449;

DT 28-FEB-2000 (first entry)

XX Mouse ESE21 protein sequence.

XX Mouse; murine; ESE2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.

XX Mus sp.

XX WO955728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-CA00375.

XX 27-APR-1998; 98CA-2230201.

XX 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI; 2000-052802/04.

DR

DR N-PSDB; AA239024, AA239025.

XX New nucleic acid encoding ESE1 and 2 proteins, involved in regulation of endocytosis, used e.g. for treating cancer or preventing viral infection

XX Claim 14; Page 62-63; 99pp; English.

XX The present invention specifically describes mammalian ESE1 and 2 proteins (I) and their splice variants (Ese - EH-domain and SH3-domain regulator of endocytosis). (I) are involved in regulation of clathrin-mediated endocytosis (as a complex with Esp15 protein), vesicular trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists, mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab); sequences, antisense to the (I) polynucleotide; agents that downregulate expression of ESE genes or antagonists of an ESE binding partner are used to treat diseases associated with undesirable endocytosis and resulting changes in cellular function. Particularly overexpression of ESE is used to block clathrin-mediated endocytosis in vivo or in cell cultures, while administration of (I) is used to promote endocytosis of selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of cells that can be stimulated to proliferate by a growth factor receptor; and similar compounds (also inactive ESE mutants) can be used to prevent viral infection. Endocytosis may also be regulated, in vivo or in cell cultures, by forming an ESE-Esp15 complex, then binding dynamin to the complex. Generally conditions that can be treated include cancer; abnormal cell division or migration; viral infection; or abnormal receptor signalling, tissue development or synaptic transmission. The present sequence represents mouse ESE1 protein sequence.

XX Sequence 1715 AA;

Query Match 15.9%; Score 163; DB 21; Length 1715;

Best Local Similarity 20.4%; Pred. No. 6.5e-08;

Matches 39; Conservative 52; Mismatches 94; Indels 6; Gaps 2;

QY 4 EVSTTERTYLKLEVITSWFQSTVSKEDAMPEALKSLIFPNFPELKHFTNFKIEQR 63

DB 1238 elivteenyvndqlqvteifqkplteselltekevamifvfwkelmcnliklralrvr- 1296

QY 64 ALWEGSRNAQIRDYQIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFC 123

DB 1297 -----kkmgekmpvkmldiisaiqlpmpqyirfscqlngaaliqqktdaepdfekvk 1352

QY 124 DFELQKVY-LPLNTFLRLPLRLHMLHYKQVLERLCKHPPSHADFRDRAALAEITEMVA 182

DB 1353 rlamdprckgplssflkpmqtrityrplliknilentpenhpdhshlkhaeaelcs 1412

QY 183 QLHGTWIKMEN 193

DB 1413 qvnegvreken 1423

RESULT 11

AA36516

ID AAB36516 standard; Protein; 844 AA.

XX AAB36516;

XX 06-MAR-2001 (first entry)

XX Candida albicans CaCDC24 protein sequence.

XX Candida albicans; yeast pathogen; identification; fungal; antifungal;

XX CaKRE5; CaALR1; CaCDC24; diagnosis; fungicide; fungal infection.

XX Candida albicans.

XX WO200068420-A2.

XX 16-NOV-2000.

XX


```
CC produce recombinant TRIO polypeptides, e.g. for antibody production
CC or screening for modulators of activity. Labelled antibodies that
CC bind TRIO specifically can be used to detect/quantify TRIO activity,
CC particularly for diagnosis and phenotyping of neoplastic or
CC hyperplastic disease. Modulators of TRIO activity (e.g. antisense
CC nucleic acids, antibodies, peptides or mimetics) can be used to
CC reorganise the actin cytoskeleton (claimed), e.g. in cases of wound
CC healing and/or tumour metastasis, to treat an oncogene (claimed),
CC or more generally to control growth, differentiation, migration
CC and/or survival of cells, e.g. regulation of the immune response to
CC infection, treatment of impaired immune response (as in chronic
CC granulomatous disease), control of apoptosis in cancer therapy, and
CC treatment of degenerative diseases (e.g. Parkinson's, Alzheimer's
CC or Huntington's, amyotrophic lateral sclerosis, gastric ulcers,
CC Wilms' tumour etc.). Transgenic animals can be used as models to
CC characterise TRIO genes and proteins.
XX
SQ Sequence 2861 AA;
Query Match 13.5%; Score 138; DB 18; Length 2861;
Best Local Similarity 22.5%; Pred. No. 5.5e-05;
Matches 41; Conservative 47; Mismatches 64; Indels 30; Gaps 7;
QY 1 IAEVSTERTYTKDLVITSWFQSTVSKEDAMPEALK---SLIFNPEPLHKFHTN-FL 56
Db 1914 vlqelveterdyvrdlgvvygyma-lmkedgvpdmkgdkivfgnhiqydwdrdfl 1972
QY 57 KELEORLALWEGRSNAQIRDYQIRIGDVMKLNIOGMKHLAAHLWKHSEALENGIKS-- 114
Db 1973 gelekl-----edpeklsifvkherrl-----hmy-----laycnpkpkseh 2011
QY 115 --SRLENFCRDFELQKVCVLTNTLLRPLRLMHYKOVLERLCKRHPPSHADRCRA 172
Db 2012 ivseyidtfedkqrlghrlqtdllikpvqimkyqlllkflkyskksldtseler 2071
QY 173 AL 174
Db 2072 av 2073
RESULT 15
AAW64468
ID AAW64468 standard; Protein; 1461 AA.
XX
AC AAW64468;
XX
DT 16-OCT-1998 (first entry)
XX
DE Human secreted protein from clone CW420_2.
XX
KW Secreted protein; nutrition; cytokine; cell proliferation; activin;
KW differentiation; immune system; stimulator; suppressor; tissue growth;
KW haematopoiesis regulation; inhibin; chemotactic; chemokinetic; cadherin;
KW haemostatic; thrombolysis; receptor; ligand; anti-inflammatory.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 364 /label= unknown
FT Misc-difference 369 /label= unknown
FT Misc-difference 1433 /label= unknown
XX
PN WO9830589-A2.
XX
PD 16-JUL-1998.
XX
PF 20-DEC-1997; 97WO-US23506.
XX
PR 18-DEC-1997; 97US-0993228.
```

```
PR 10-JAN-1997; 97US-0781225.
XX (GEMY ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
XX WPI; 1998-413682/35.
DR N-PSDB; AAV46316.
XX
XX New isolated nucleic acids and secreted proteins - isolated from
PT human foetal kidney, adult placenta, adult colon, adult testes,
PT foetal brain and adult brain cDNA libraries
XX
XX Claim 24a; Page 75-80; 122pp; English.
XX
XX This sequence represents a novel secreted protein from clone CW420_2
CC isolated from a human fetal brain cDNA library. This novel protein may
CC have biological activity, e.g. nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity or other activities.
XX
SQ Sequence 1461 AA;
Query Match 12.8%; Score 131; DB 19; Length 1461;
Best Local Similarity 22.7%; Pred. No. 0.00012;
Matches 47; Conservative 43; Mismatches 97; Indels 20; Gaps 8;
QY 1 IAEVSTERTYTKDLVITSWFQSTVSKEDAMPEALKSLIFNPEPLHKFHTNFKLEIE 60
Db 623 vinelfyterahrvtlkvdqvfyqrvsregilspselrkifsnledilqlhig----in 678
QY 61 ORLALWEGRSNAQIRDYQIRIGDVMKLNIOG-----MKHLAAHLWKHSE-ALEALENGIKS 114
Db 679 eqmkavrkknetsvid--qigedlltwfsgpgeeklkhaaatfcsnqpfalemiksrqkk 736
QY 115 SRLENFCRDFELQKVC-YLPLNTLLRPLRLMHYKQVLERLCKH--HPPSHADFR--- 168
Db 737 dsrfgtfgvqdaesnpicrrllqldkdiptgmqrtrkypllldniakytewpterekvkaa 796
QY 169 -DCRAALAEITEMVAQLHGTMIKWFNF 194
Db 797 dhcrqilnyngavkveenqk-rledy 822
```

Search completed: December 6, 2001, 08:51:43
Job time: 335 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:55:09 ; Search time 36.72 Seconds
(without alignments)
118.890 Million cell updates/sec

Title: US-09-555-342A-2_COPY_544_737

Perfect score: 1023

Sequence: 1 IAKEVSTERTYTKDLEVT.....AEITEMVAQLHGTMIKMF 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	13.5	761	4	US-09-625-188-14
2	138	13.5	2860	2	US-08-826-267-2
3	131	12.8	1461	2	US-08-993-228-10
4	120	11.7	1244	3	US-09-356-952-7
5	118	11.5	477	1	US-08-191-338A-2
6	94.5	9.2	912	4	US-08-943-768-2
7	86.5	8.5	626	4	US-09-155-770-7
8	80	7.8	844	1	US-07-646-537B-2
9	78.5	7.7	1724	2	US-08-477-451-15
10	77.5	7.6	583	2	US-08-616-392C-4
11	77.5	7.6	645	1	US-08-785-430-2
12	77.5	7.6	645	2	US-08-996-800-2
13	76.5	7.5	2154	2	US-08-841-349-4
14	76	7.4	534	2	US-08-878-563A-1
15	76	7.4	534	4	US-09-270-117-1
16	75.5	7.4	748	3	US-08-904-871-6
17	75.5	7.4	748	3	US-08-904-871-13
18	75	7.3	521	2	US-08-878-563A-3
19	75	7.3	521	4	US-09-270-117-3
20	74	7.2	584	2	US-08-415-593-41
21	74	7.2	1143	2	US-08-310-912A-108
22	74	7.2	1143	5	PCT-US95-04589-108
23	74	7.2	1144	5	PCT-US95-04589-108
24	74	7.2	1144	5	US-08-261-663A-2
25	74	7.2	1144	4	US-08-261-663A-4
26	74	7.2	1144	3	US-08-930-996A-9
27	74	7.2	1144	5	PCT-US95-07754A-2

```
28 74 7.2 1144 5 PCT-US95-07754A-4
29 73.5 7.2 487 2 US-08-724-394A-7
30 72.5 7.1 871 2 US-08-775-009-34
31 72.5 7.1 871 2 US-08-775-009-35
32 71.5 7.0 732 1 US-08-481-626-2
33 71.5 7.0 732 4 US-08-989-299-4
34 71.5 7.0 1306 4 US-08-989-299-7
35 71 6.9 326 2 US-08-306-511A-10
36 71 6.9 326 2 US-08-893-274-10
37 71 6.9 326 3 US-08-581-918A-10
38 71 6.9 326 4 US-08-346-147B-10
39 71 6.9 326 4 US-08-822-936-10
40 71 6.9 326 5 PCT-US95-04636-10
41 71 6.9 1911 1 US-08-348-006B-5
42 71 6.9 1911 2 US-08-800-825A-5
43 71 6.9 1911 4 US-09-158-657-5
44 71 6.9 1911 5 PCT-US94-10166-5
45 70.5 6.9 297 2 US-09-027-013-3
```

ALIGNMENTS

```
RESULT 1
US-09-625-188-14
; Sequence 14, Application US/09625188
; Patent No. 6307037
; GENERAL INFORMATION:
; APPLICANT: No. 6307037artis AG
; TITLE OF INVENTION: Fungal Target Genes and Methods
; FILE REFERENCE: PB/5-31285P1
; CURRENT APPLICATION NUMBER: US/09/625,188
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 761
; TYPE: PRT
; ORGANISM: Ashbya gossypii
US-09-625-188-14
```

Query Match 13.5%; Score 138; DB 4; Length 761;

Best Local Similarity 26.2%; Pred. No. 1.5e-07;

Matches 56; Conservative 37; Mismatches 83; Indels 38; Gaps 9;

```
QY 1 IAKEVSTERTYTKDLEVTSTVSTVSKEDAMPEALKSLIFNFEPLHFKHFTNFKIE 60
DB 267 IVKEFVETERYVHDDEVLSKYRQQLLENIISSEEL-YMLFFNLNEIIDQRRFLVALE 325
QY 61 QRLALMEGRSNAQI-RDYQRIGDVMLKNIQGMKHLAAHLKMKHSEALENGIKSS-RRL 118
DB 326 -----INGQVPAQAQRIGALFMHS----KHFFKLYEPWSIGQNAAINFISSSPDKM 372
QY 119 ENCFRDFEQKVCYLPNTFLPLRLHRLMHYKQVLERLCK--HPPSHADFRCRAAL-- 174
DB 373 QS--QEFVIGN--KMELQSFLLKPVQRLCRYPLLLKDLKLSVKTKSDVDTKQLTALEI 428
QY 175 -----AEITEMVAQLHGTMIKMF 194
DB 429 SKSIARSINENQRTTENHEVVKLYGRVNNKGY 462
```

RESULT 2

```
US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: No. 5994070el TRIO Molecules and Uses Related Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
```

1

US-09-356-952-7

Query Match 11.7%; Score 120; DB 3; Length 1244;
Best Local Similarity 22.1%; Pred. No. 4.2e-05;
Matches 44; Conservative 33; Mismatches 90; Indels 32; Gaps 5;
QY 9 ERTYKDLVITSWF-----QSTVSKEDAMPEALKSLIFPNFELPKHFTNFKETEORL 63
DB 252 EAEYVQQLHILVNFRLPLRMAASSKPPITHDDVSSIFLNSSETIMFLHQIFVQGLKARI 311
QY 64 ALWEGRSNAIRD-----XORIGDVMLKNTQGMKHLAAHLWKHSEALEALENGIKSS 115
DB 312 ASWPTLVLDLFDLLPMNITQE-----FVRNHQVSLQILAH-----CKON 353
QY 116 RRLNFCRDFELQKVC-YLPLNTFLRLPLRLHMLHYKQVLERLCKHHPPSHADPRDRAAL 174
DB 354 RQFDKLLKQYEAKPDCEERTLETFELTYPFQIPRYTLTLLHLLAHTPHEHVERNSLDYAK 413
QY 175 AEITEMVAOLHGTMIKEN 193
DB 414 SKLELSRVMHDEVSTEN 432

RESULT 5

US-08-191-338A-2
; Sequence 2, Application US/08191338A
; Patent No. 5763164

; GENERAL INFORMATION:
; APPLICANT: Calenoff, Emanuel
; TITLE OF INVENTION: Immunogenic Cancer Proteins and Peptides
; TITLE OF INVENTION: and Method of Use
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:

; ADDRESSER: Arnold, White & Durkee
; STREET: 321 No. 5763164th Clark Street, Suite 800
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60610

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/191,338A
; FILING DATE: 12-JAN-1994

; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5763164thrup, Thomas E.
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: nwn:002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-744-0090
; TELEFAX: 312-755-4489

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-191-338A-2

Query Match 11.5%; Score 118; DB 1; Length 477;
Best Local Similarity 21.0%; Pred. No. 1.7e-05;
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;

QY 1 IAKEVSTTERTYKDLVITSWFQSTVSKEDAMPEAL---KSLIFPNFELPKHFTN 54
DB 52 VLNELIQTERVYVRELYTVLLGYRAEMDNPEMFDLMPPLLRNKKDILFCGMAEIVEFHND 111

QY 55 FLKEIEQRLALWEGRSNAQIRDYQIRIGDVMLKNTQGMKHLAAH-----LW-KHSEA 104
DB 112 IF-----LSSLENCAP-----ERVGPCFLERKDDDFQYAKYQCNKPRSETIWRKYSEC 161
QY 105 LEALENGIKSSRRLENFCRDFELQKVCYLPNTFLRLPLRLHMLHYKQVLERLCKHHPPSH 164
DB 162 AFQECQCRKLKRLR-----LDSYLLKPVORITKYQLLLKELLKYS----- 202
QY 165 ADPRDC-----RAALAEITEMVAOLHGTM 188
DB 203 ---KCEGSALLKALDAMDLLKSYNSDM 229

RESULT 6

US-08-943-768-2
; Sequence 2, Application US/08943768C
; Patent No. 6238881

; GENERAL INFORMATION:

; APPLICANT: Hart, Matthew J.

; TITLE OF INVENTION: No. 6238881el Nucleic Acids and Polypeptides Related to a
; TITLE OF INVENTION: Guanine Exchange Factor of RHO GTPase

; FILE REFERENCE: 1023-US

; CURRENT APPLICATION NUMBER: US/08/943,768C

; CURRENT FILING DATE: 1997-10-06

; EARLIER APPLICATION NUMBER: 60/029,979

; EARLIER FILING DATE: 1996-11-06

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 912

; TYPE: PRT

; ORGANISM: Human p115 GEF-Rho

US-08-943-768-2

Query Match 9.2%; Score 94.5; DB 4; Length 912;

Best Local Similarity 22.0%; Pred. No. 0.027;

Matches 47; Conservative 45; Mismatches 87; Indels 35; Gaps 10;

QY 1 IAKEVSTTERTYKDLVITSWFQSTVSKEDAMP-EALKSLIFPNFELPKHFTNFKEI 59

DB 420 VISELLVTEAAHVRLVLDLFFQPMACCLFPPLBELQN-IPPSLDELIEVLSLFDRL 478

QY 60 EQRLALWEGRSNAQIRDY--QRIGDVMLKNTQGMKHLAAHL-----KHSEALEA 107

DB 479 MKR-----RQESGYLIEEIGDVLARFDG----AEGSWFKITSSRFSRQSFALQ 525

QY 108 LENGIKSSRRLENFCRDFELQKVC-YLPLNTFLRLPLRLHMLHYKQVLERLCKH-HPPSHA 165

DB 526 LKAKQKDPKFCFAFVQEAESRPRRLQLKDMIPTQMRLTKYPLLQSIGQNTSEPTER 585

QY 166 DFRD-----CRAALAEITEMVAOLHGTMIKENF 194

DB 586 EKVELAAECCREILHHVNOAVRDME-DLLRLKDY 618

RESULT 7

US-09-155-770-7

; Sequence 7, Application US/09155770A

; Patent No. 6300484

; GENERAL INFORMATION:

; APPLICANT: Duhl, David

; TITLE OF INVENTION: DNA ENCODING DP-75 AND A PROCESS FOR ITS USE

; FILE REFERENCE: 200130.418

; CURRENT APPLICATION NUMBER: US/09/155,770A

; CURRENT FILING DATE: 1998-09-30

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 7

; LENGTH: 626

; TYPE: PRT

; ORGANISM: Homo sapien

```

Query Match      7.88; Score 80; DB 1; Length 844;
Best Local Similarity 33.9%; Pred. No. 1.2;
Matches 21; Conservative 11; Mismatches 28; Indels 2; Gaps 2;

QY 3 KEVSTTERTYKOLEVITSWFSQSVKSDAMPEAKSLIFPNFPLAKHFNFLKEIQOR 62
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 REIQOTEEKYTDTLGSIQQHFMKPLQR-FUKPODMET-IFVNIIEELFSVHTHFLKELKDA 257

QY 63 LA 64

```

RESULT 10
US-08-616-392C-4
: Sequence 4, Application US/08616392C
: Patent No. 5998165
: GENERAL INFORMATION:
: APPLICANT: Goold, Richard D.
: APPLICANT: Akerblom, Ingrid E.
: APPLICANT: Seilhamer, Jeffrey
: APPLICANT: Coleman, Roger
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDE
: TITLE OF INVENTION: AND PAMCIB A

Query Match 7.6%; Score 77.5; DB 1; Length 645;

Matches	51;	Conservative	32;	Mismatches	85;	Indels	77;	Gaps	10;
BEST LOCAL SIMILARITY 20.8%, FREQ: NO. 1.0;									

QY 2 AKEVSTTERTYLYKDLVITSNFQSTVSKEDAMPEALKSLIFPNFEPLHKFHTNFKLEIQ 61

Query Match 7.6%; Score 77.5; DB 2; Length 583;

Best Local Similarity 23.0%; Pred. No. 1.4;
Matches 37; Conservative 26; Mismatches 69; Indels 29; Gaps 6;

Qy 15 DLEVITSWFQSTVSKEDAMPEALKSLTFPNEFPLHKFTNFKIEQRLALWEGRSNAQI 74

Db 98 DSQVLQEARPPLVSADDEIYSTSKAFIGIPIYKPPKKN-----EGRNEAHV 145

QY 75 RD--YQRIGDVMLKNIQGMK-HLAAHLWKHSEALELENGIKSSRRLENFCRDFELQKVC 131

Db 146 LNGINDRGQKEKQKFNSEKSEIDNELFQFYKEIEELE--KEKDGFFENSCKESEPSEQ 202

Qy 132 YLPL----NTFLLRPLRLMHYKQVLERLCKHHPPSHADF 168

db 203 EVPEVEGHNNGLIKRP-----DEEKKDLSNKAMPSCDYO 236

RESULT 11

US-08-785-430-2

; Sequence 2, Application No. 5795757

GENERAL INFORMATION:

APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth

TITLE OF INVENTION: No. 5795757e1 tRNA Synthetase

; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road
CITY: King of Prussia

City: KING
; STATE: PA

; COUNTRY: USA
; ZTP: 19406-0939; ;
ZIF. 19400 0933
COMPUTER READABLE FORM:

```

; . MEDIUM TYPE: Diskette

```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,800
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING APPLICATION NUMBER: 08/785,430
; FILING DATE: 17-JAN-1997
; APPLICATION NUMBER: 9601096.2
; FILING DATE: 19-JAN-1996
; APPLICATION NUMBER: 9622617.0
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31354-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-996-800-2

Query Match 7.6%; Score 77.5; DB 2; Length 645;
Best Local Similarity 20.8%; Pred. No. 1.6;
Matches 51; Conservative 32; Mismatches 85; Indels 77; Gaps 10;

QY 2 AKEVSTERTYRKDLEVITSWFQSTVSKEDAMPEALKSLIFFNPEPLHKFHTNFKIEQ 61
Db 146 AKELFSDNDEYKLELIDAIPEVDENVTLYSQGDTDLGRGVHVPSTAKIKEP-----KLLST 200
QY 62 RLALWEGRSNAQRDQYRIGDVMLKNTQ-----MKHLAHLWKHSEALEA----- 107
Db 201 AGAYWRGDSN-----NKMLQRIYGTAFFDKKELKAHLQMLEERKDRHRKIGKEL 250
QY 108 -----LENGIKSSRRLENFCRDFELQ---KVCYLPL--NTFLLRPLHRL 146
Db 251 ELEFNSOLVAGLPLWLPNGATIRREIERIYVDKEYSMGVYDHYVTPVLNVLDLYKTSQHW 310
QY 147 MHYKQ-----VLERL-CKHHPPSHAD-----FRDCRAALAEITEMVAQLHGT 187
Db 311 DHYQEDMFPQMQLDETESMVLRPWNCPPHMMIYANKPHSYRELPIRIAEI-----GT 362
QY 188 MIKWE 192
Db 363 MHRYE 367

RESULT 13
US-08-841-349-4
; Sequence 4, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO4470US0
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2154
; TYPE: PRT
; ORGANISM: Mus musculus

```

```

US-08-841-349-4

Query Match 7.5%; Score 76.5; DB 2; Length 2154;
Best Local Similarity 24.1%; Pred. No. 1.3;
Matches 51; Conservative 29; Mismatches 61; Indels 71; Gaps 14;

QY 3 KEVSTTE----RTYKLDLEVITSWF---QSTVSKEDAMPEALKSLIFFNPEPLHKFHTNF 55
Db 1041 REASLGESKLOQFLRDLDDFQSWLSRTQTAIASED--MPNTLTEA-----EKLTTQHENI 1094
QY 56 LKEIQRLALWEGRSNAQRDQYRIGDVMLKNTQGMKHLAAHLWKHSEALEALENGIKSS 115
Db 1095 KNEIDN---YE-----EDYQKMRDMGEMVTQGGTD-AQYMFRLR-QRLQALDTG---- 1137
QY 116 RRLNEFCROFELQKVCYLPLNTLTPRLPLHLMHYKVLERLCKHHPPSHADF-RDCRAAL 174
Db 1138 -----WNLHKMMWENRNL--LSQSH--AYOOFLRDTKQAE 1169
QY 175 AEI-----TEMVAQLHG---TMIKMENF 194
Db 1170 AFLNNQYVLAHTEMTPTTLEGAAAIKKQEDF 1201

RESULT 14
US-08-878-563A-1
; Sequence 1, Application US/08878563A
; Patent No. 5891674
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,563A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0323 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 534 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT04
; CLONE: 918158
; US-08-878-563A-1

```

Query Match 7.4%; Score 76; DB 2; Length 534;
Best Local Similarity 24.8%; Pred. No. 1.8;

Job time: 293 sec

Qy	46	EPLHKFTNFKETEQLALWEGSRNAQIRDYQ-----RTGDV-----MLKNIQGM	91
		: : : : :	
Db	92	EMLKSFNELLTQLEQVELDLSRYLSAALKKYYQTEQRSGDALDKQAEELKLRKKSOGS	151
		: : : : :	
Qy	92	KHLAAHLWKHSEALEALENGKISRSRLNFCRD-----FELQKVCYL	133
		: : : : :	
Db	152	KNPKQYSDKRELQIDATSN--KQGELENVSDGYKALTAECCRRFCFL	197
		: : : : :	

RESULT 15
 US-09-270-117-1
 ; Sequence 1, Application US/09270117
 ; Patent No. 626550
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Shah, Purvi
 ; TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/270,117
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/878,563

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0323 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT04
CLONE: 918158
US-09-370-117-1

```

Query Match          7.4%; Score 76; DB 4; Length 534;
Best Local Similarity 24.8%; Pred. No. 1.8;
Matches 27; Conservative 18; Mismatches 40; Indels 24; Gaps

Qy 46 EPLHKFTNPLKEIEORLALWEGRSNAQINDYQ-----RIGDV-----MLKNIQGM 91
| | | | | : | | | : | | | | | | | | | | | |
Db 92 EMLKSPHNELLTQEQVELDSRYLSAALKKYQTEQRSKGDALDCAELKLRKKSQGS 151
| | | | | : | | | : | | | | | | | | | | | |

Qy 92 KHLAAHLWKHSEALEALENCIKSRRLNFCRD-----FELQKVCYL 133
| : : : : | | | | | | | | | | | | | | | |
Db 152 KNQPKYSDKELQYIDAIN---KQGELENYVSDGYKALTETECRRCFEL 197
| : : : : | | | | | | | | | | | | | | | |

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:52:37 ; Search time 44.17 Seconds
(without alignments)
334.568 Million cell updates/sec

Title: US-09-555-342A-2_COPY_544_737
Perfect score: 1023
Sequence: 1 IAKEVSTTERTYKDLVIT.....AEITEMVAQLHGTMIKMENF 194

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR:68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	100.0	1045	2 JC5795	CDEP protein - hum
2	190.5	18.6	875	2 T19678	hypothetical prote
3	171	16.7	496	2 T46356	hypothetical prote
4	166.5	16.3	1335	2 T18289	racGAP protein - s
5	157.5	15.4	1737	2 A92235	unconventional myo
6	154	15.1	961	2 A5380	faciogenital dyspl
7	129.5	12.7	573	2 I49342	85K SH3 domain-con
8	123	12.0	646	2 JC5583	hypothetical oncop
9	122	11.9	1189	2 T42726	guanine nucleotide
10	121.5	11.9	1158	2 T50454	probable rhoL GDP-
11	120	11.7	1244	2 S29083	guanine-nucleotide
12	119	11.6	552	2 T34428	hypothetical prote
13	119	11.6	2488	2 T42739	guanine nucleotide
14	118	11.5	478	4 TVHUBD	transforming prote
15	118	11.5	925	1 TVHUBD	transforming prote
16	117.5	11.5	738	2 S32372	transforming prote
17	117.5	11.5	2044	2 T13704	still life protein
18	117.5	11.5	2064	2 T13707	still life protein
19	116	11.3	872	2 T37789	Scd1 protein - fis
20	116	11.3	1260	2 S28407	guanine nucleotide
21	114.5	11.2	1334	2 T41524	rhoL gdp-ntp excha
22	114	11.1	567	2 S10138	mcf2 protein - hum
23	114	11.1	736	2 A27477	cell division cont
24	113.5	11.1	620	2 T16657	hypothetical prote
25	112	10.9	685	2 JC6331	rho-type guanine e
26	111	10.9	872	2 S51620	Ost oncogene - rat
27	111	10.9	1275	2 A38985	nucleotide exchang
28	110	10.8	677	2 T15242	hypothetical prote
29	110	10.6	1275	2 T41523	hypothetical rhoL

30 106.5 10.4 278 2 A60195 transforming prote
31 105.5 10.3 1693 2 T30867 Rho-guanine nucleo
32 105 10.3 914 2 T17233 hypothetical prote
33 105 10.3 1271 1 TVHUBR bcr (breakpoint cl
34 104 10.2 1591 2 A54146 invasion-inducing
35 103.5 10.1 860 2 T15778 hypothetical prote
36 103.5 10.1 919 2 T21663 hypothetical prote
37 103 10.1 878 2 T51940 gene VAV2 protein
38 101 9.9 462 2 T25544 hypothetical prote
39 101 9.9 519 2 T38402 guanine nucleotide
40 98 9.6 460 2 G01210 guanine nucleotide
41 98 9.6 1899 2 T32732 PAM C-terminal int
42 98 9.6 1919 2 T42098 PAM interacting pr
43 97.5 9.5 862 2 T34342 hypothetical prote
44 96.5 9.4 2137 1 SJHUB spectrin beta chai
45 95.5 9.3 693 1 TVHUA2 bcr (breakpoint cl

ALIGNMENTS

RESULT 1
JC5795
CDEP protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C:Accession: JC5795
R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.
Biochem. Biophys. Res. Commun. 241, 369-375, 1997
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain
ng factors.
A:Reference number: JC5795; MUID:98086358
A:Accession: JC5795
A:Molecule type: mRNA
A:Residues: 1-1045 <KOY>
A:Cross-references: DBJ:AB008430
C:Comment: this protein is involved in the adhesion, proliferation, and differentiat
ion.
F:1-374/Domain: ezrin-like #status predicted <E2R>
F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>
F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 100.0%; Score 1023; DB 2; Length 1045;
Best Local Similarity 100.0%; Pred. No. 9.3e-80;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAKEVSTTERTYKDLVITSNFQSTVSKEDAMPEALKSLIFPNFPLHKEHTNFKLEIE 60
|||||
Db 544 IAKEVSTTERTYKDLVITSNFQSTVSKEDAMPEALKSLIFPNFPLHKEHTNFKLEIE 603
|||||

QY 61 QRLALWEGRSNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSEALEENGKSSRRLEN 120
|||||
Db 604 QRLALWEGRSNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSEALEENGKSSRRLEN 663
|||||

QY 121 FCRDFELQVCVYLPLNTFLLRLHLMHYKQVLERLCKHHPHSHADFRCRAALAEITEM 180
|||||
Db 664 FCRDFELQVCVYLPLNTFLLRLHLMHYKQVLERLCKHHPHSHADFRCRAALAEITEM 723
|||||

QY 181 VAQLHGTMIKMENF 194
|||||

Db 724 VAQLHGTMIKMENF 737
|||||

RESULT 2
T19678
hypothetical protein C33D9.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19678
R:Lloyd, C.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19162

A:Accession: T19678
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-875 <NLS>
A:Cross-references: EMBL:Z68159; PIDN:CAA92282.1; GSPDB:GN00022; CESP:C33D9.1
A:Experimental source: clone C33D9
C:Genetics:
A:Gene: CESP:C33D9.1
A:Map position: 4
A:Introns: 14/1; 74/3; 129/3; 171/3; 198/1; 241/2; 269/3; 352/3; 403/2; 515/3; 595/3; 65/3

Query Match 18.6%; Score 190.5; DB 2; Length 875;
Best Local Similarity 25.6%; Pred. No. 1.7e-08;
Matches 50; Conservative 46; Mismatches 88; Indels 11; Gaps 4;

QY 2 AKEVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNPEPLHKFH-TNFKETE 60
Db 333 AKELVDSEQRVYDKLLGDTFRNRIKEELITNDITRLANVSSLYQFHNTHTFLPQLL 392
QY 61 ORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRLEN 120
Db 393 ESTRDW-----HTTKRIANVVRKQAPFLKMYSETTNNIDRATKLFEE-LKKKKKPAD 443
QY 121 FCRDFELQKVCY-LPLNTFLRLPLRLMHYKQVLERLCKRHPPSHADFRDCRAALAEITE 179
Db 444 VKIEIEKQAECEGLPLGHHLICPVQVRMYQLLLQYKHLQPSDVFDDTTVALELVQL 503
QY 180 MVAQLHGTMIKMFN 194
Db 504 AAHAHEMMKKLDRF 518

RESULT 3
T46356
hypothetical protein DKFp434G2016.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46356
R:Koehrer, K.; Beyer, A.; Meves, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23037
A:Accession: T46356
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-496 <AAA>
A:Cross-references: EMBL:AL137289
A:Experimental source: adult testis; clone DKFp434G2016
C:Genetics:
A:Note: DKFp434G2016.1

Query Match 16.7%; Score 171; DB 2; Length 496;
Best Local Similarity 26.0%; Pred. No. 4e-07;
Matches 51; Conservative 37; Mismatches 96; Indels 12; Gaps 4;

QY 1 IAKEVSTTERTYKDL-EVITSWFQSTVSKEDAMPEALKSLIFPNPEPLHKFHTNFKEI 59
Db 94 VINEILSTERYIKHLRDICEGVVRCRRADWFSEQLRTIFGNIEDIYRCQKAFVKAL 153
QY 60 EORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRLE 119
Db 154 EQRF-----NRERPHLSLGACFLHQADFOIYSEYCNHNPACVLSRLTKLSKYVY 206
QY 120 NF--CRDFELQKVCYLPNTFLRLPLRLMHYKQVLERLCKRHPPSHADFRDCRAALAEI 177
Db 207 FFPACR--LQKMDISLSDGFLTTPVQCKYPLQAEILLKTYHPQHRDFKDVAAALHAM 264
QY 178 TENVAQLHGTMIKMFN 193
Db 265 KNVAQLINERKRLEN 280

RESULT 4
T18289
racGAP protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T18289
R:Ludbrook, S.B.; Eccleston, J.; Strom, M.
J. Biol. Chem. 272, 15682-15685, 1997
A:Title: Cloning of a rhoGAP homolog from Dictyostelium discoideum.
A:Reference number: Z18859; MUID:97332648
A:Accession: T18289
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1335 <LUD>
A:Cross-references: EMBL:Y10159; NID:g2190354; PIDN:CAA71241.1; PID:g2190355
C:Genetics:
A:Gene: racGAP
A:Introns: 113/1; 205/3

Query Match 16.3%; Score 166.5; DB 2; Length 1335;
Best Local Similarity 27.6%; Pred. No. 3.2e-06;
Matches 59; Conservative 34; Mismatches 74; Indels 47; Gaps 9;

QY 2 AKEVSTTERTYKDLVITSWFQSTVSK-----EDAMPEALKSLIFPNPEPLHKFH 52
Db 795 AEETLTTEQYVYKQLTVVYDNFIEPYKTKQKCHGLTGESEFMD-----IFNCLEVILSSH 848
QY 53 -TNFLUKEIEORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENG 111
Db 849 KTNLLKPIEDRLVMSK-----PMGDIFLNNTSFIK-----LYKH-----YVNNY 890
QY 112 IKSSRLENFCRD---FEL-----QKVCYLPNTFLRLPLRLMHYKQVLERLCKH 159
Db 891 DRSIRTL-NQCKEYDGFYTMASLDYSENLSLESFLVLPQIQLPRYVMLLQDLKY 949
QY 160 HPPSHADFRDCRAALAEITE MVAQLHGTMIKMFN 193
Db 950 TANDHEDFNQCEALSTIKDLTESINTKKSEEDN 983

RESULT 5
A59235
unconventional myosin heavy chain Myom - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
C:Accession: A59235
R:Geissler, H.; Schwarz, E.C.; Soldati, T.
submitted to GenBank, September 1998
A:Description: Identification of two novel and highly divergent myosins in Dictyostelium
A:Reference number: A59235
A:Accession: A59235
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1737 <GEI>
A:Cross-references: GB:AF090533; NID:g5714395; PIDN:AAD47903.1; PID:g5714396
A:Experimental source: strain AX2
C:Genetics:
A:Gene: myom
A:Map position: 6, aldB-caba2
A:Superfamily: myosin motor domain homology
F:62-874/Domain: myosin motor domain homology #status atypical <NMO>

Query Match 15.4%; Score 157.5; DB 2; Length 1737;
Best Local Similarity 25.9%; Pred. No. 2.6e-05;
Matches 49; Conservative 37; Mismatches 82; Indels 21; Gaps 7;

QY 1 IAKEVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNPEPLHKFHTNFKETE 60
Db 1393 IINELIETERYDVRDLNIVVEFLNPIREKQLLSAKDINSLSFNILFSINNVLKALE 1452
QY 61 Q-RLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALEAL-ENGIKSS--R 116

Db 1453 KKDPLCENIS-----VCQTFELMSHYLKMWTYCSNOONALKILEEKIKNOPPR 1503
QY 117 RLNEFCRDFELQKVCY-LPLNTFLRLRLMHYKQVLERLCKHHPPSHADFRDCRAALA 175
Db 1504 EYLEFCMN--DSVCRGLPLNSFIKPKVQRICKYPLLIKETIKFTPNDDPD---KPALE 1556
QY 176 EITEMVAQL 184
Db 1557 EYDKKISDI 1565
RESULT 6
A55380
faciogenital dysplasia-associated protein FGD1 - human
C:Species: Homo sapiens (man)
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
C:Accession: A55380
R:Pastoris, N.G.; Cadle, A.; Logie, L.J.; Porteous, M.E.M.; Schwartz, C.E.; Stevenson, R.
Cell 79, 669-678, 1994
A:Title: Isolation and characterization of the faciogenital dysplasia (Aarskog-Scott syn-
A:Reference number: A55380; MUID:95042764
A:Accession: A55380
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-961 <PAS>
A:Cross-references: GB:U11690; NID:g595424; PID:g595425
C:Superfamily: CDC24 homology; pleckstrin repeat homology
F:373-561/Domain: CDC24 homology <CD24>
Query Match 15.1%; Score 154; DB 2; Length 961;
Best Local Similarity 23.7%; Pred. No. 2.5e-05;
Matches 47; Conservative 41; Mismatches 96; Indels 14; Gaps 4;
QY 1 IAKEVSTTERTYKLDLEVTISWFQSTVSKE----DAMPEALKSLIFPNFELHKEHTNP- 55
Db 377 IANELLQTEKATVSRHLDDQVFCARLLEEARNRSPADVVHGFISNICSICYFQQQL 436
QY 56 LKEIEORLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSS 115
Db 437 LPELEKREWEEND-----RYPRIQDILQLAPFLKMYGEVKNFDRAVELVNTWTERS 488
QY 116 RLNEFCRDFELQKVCY-LPLNTFLRLRLMHYKQVLERLCKHHPPSHADFRDCRAAL 174
Db 489 TQFKVIIHEVQKEEACGNLTQLQHMLEPVQIRPRYELLKDYLLKPLPHGSPDSKDAQSL 548
QY 175 AEITEMVAQLHCTMIKME 192
Db 549 ELIATAAEHSNAAIRKME 566
RESULT 7
I49342
hypothetical oncogene protein lfc [imported] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Jun-2000
C:Accession: I49342
R:Whitehead, I.; Kirk, H.; Tognon, C.; Trigo-Gonzalez, G.; Kay, R.
J. Biol. Chem. 270, 18386-18395, 1995
A:Title: Expression cloning of lfc, a novel oncogene with structural similarities to gua-
A:Reference number: A57387; MUID:95355462
A:Accession: I49342
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-573 <RES>
A:Cross-references: EMBL:U28495; NID:g902022; PIDN:AAC52234.1; PID:g902023
C:Genetics:
A:Gene: lfc
C:Superfamily: pleckstrin repeat homology
F:472-570/Domain: pleckstrin repeat homology <PLK>

Query Match 12.7%; Score 129.5; DB 2; Length 573;
Best Local Similarity 21.4%; Pred. No. 0.0017;
Matches 44; Conservative 50; Mismatches 85; Indels 27; Gaps 7;
QY 4 EYSTTERTYKLDLEVTISWFQSTVSKEADAM-PEALKSLIFPNFELHKEHTNFKELI--E 60
Db 243 ELIQTELHVRILKINLRLFTGMLQELQMEPEVVOGL-FPCVDELSDIHTRFLNQLLER 301
QY 61 QRLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKS 114
Db 302 RRQALCPGSTRNFV--IHLRGDLLISQFSGSNAEQMRKTYSEFCSRHTKALKLYKELYAR 359
QY 115 SRLENFCRDFELQKV-----CYLPLNTFLRLRLMHYKQVLERLCKHHPPSHAD 166
Db 360 DKRFQOFIRKMTSRSAVLKRGVQEC-----ILLVQRTIKTPVPLINILNONGHVEEE 412
QY 167 FRDCRAALAEITEMVAQLHCTMIKME 192
Db 413 YODLASALGLVKELLSNVDDVHELE 438
RESULT 8
JC5583
85K SH3 domain-containing proline-rich protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 16-Dec-1998
C:Accession: JC5583
R:Oh, W.K.; Yoo, J.C.; Jo, D.; Song, Y.H.; Kim, M.G.; Park, D.
Biochem. Biophys. Res. Commun. 235, 794-798, 1997
A:Title: Cloning of a SH3 domain-containing proline-rich protein, p85SPR, and its loc-
A:Reference number: JC5583; MUID:97350865
A:Accession: JC5583
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-646 <OHA>
A:Cross-references: GB:U96634
A:Experimental source: thymus
C:Comment: This protein interacts with proteins in focal complexes.
C:Superfamily: SH3 homology; CDC24 homology
F:13-60/Domain: SH3 homology <SH3>
F:93-273/Domain: CDC24 homology <CD24>
F:215-242/Domain: CDC24 homology; leucine zipper #status predicted <LZP>
F:276-279/Region: nuclear location signal
F:408-515/Region: proline-rich
F:508-511/Region: nuclear location signal

Query Match 12.0%; Score 123; DB 2; Length 646;
Best Local Similarity 23.4%; Pred. No. 0.0072;
Matches 43; Conservative 31; Mismatches 98; Indels 12; Gaps 3;
QY 1 IAKEVSTTERTYKLDLEVTISWFQSTVSKEADAMPEALKSLIFPNFELHKEHTNFKELI 60
Db 97 VLQNTLETEHEYSKELOSVLSTYLRPLQTSKLSANTSYLMGNLEEISFQOVLVQSLE 156
QY 61 QRLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSRRLE 119
Db 157 ECTKSPEAQ-----QRVGCGFSLMPQMRTLYLAYCANHPSAVSVL---TEHSEDLG 205
QY 120 NCRDFELQKVCY-LPLNTFLRLRLMHYKQVLERLCKHHPPSHADFRDCRAALAEITE 179
Db 206 EFMETKGASSPGILVLTGSLGSPFMRDLKPYTLLEKELRHMEDYHDPDRQDIQKSMATFKN 265
QY 180 MVAQ 183
Db 266 LSAQ 269

RESULT 9
T42726
guanine nucleotide release/exchange factor Ras-GRF2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Aug-2000

C:Accession: T42726
R:Fam, N.P.; Fan, W.; Zhang, L.; Chen, H.; Moran, M.F.

Mol. Cell. Biol. 17, 1396-1406, 1997

A:Title: Cloning and characterization of Ras-GRF2, a novel guanine nucleotide exchange factor

A:Reference number: 222245; MUID:97184464

A:Accession: T42726

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1189 <FAM>

A:Cross-references: EMBL:U67326; NID:gl655940; PID:gl655941; PIDN:AAC53058.1

A:Experimental source: brain

C:Genetics:

A:Map position: 13

C:Superfamily: CDC25-type guanine nucleotide exchange activator homology
F:950-1186/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 11.9%; Score 122; DB 2; Length 1189;

Best Local Similarity 23.0%; Pred. No. 0.018;

Matches 44; Conservative 34; Mismatches 97; Indels 16; Gaps 5;

QY 9 ERTYKDLVITSWF-----QSTVSKEDAMPEALKSLIFPNFELPKHFTNFKLEIQR 63

DB 255 ETEYHQLYILVNGFLRPLGMAASSKKPPINHDDVSSIFLNSETIMFLHEIFHQGLKARL 314

QY 64 ALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSSRRLENFCR 123

DB 315 ANWPTLVAD-----LFDPILLPMLNIYQEFVRN---HQYSLQVLAN-CKQNRDFDKLLK 364

QY 124 DFLQKVCY-LPLNTFLRLHRLMHYKQVLERLCKKHPPSHADFRDCRAALAEITEMVA 182

DB 365 QYENAPACEGRMLETFLTPMFIPIRYITLHLLAHTPHEHVERKSLEFAKSKLELSR 424

QY 183 QLHGTMIKMEN 193

DB 425 VMHDEYSDTEN 435

RESULT 10

T50454

probable rho1 GDP-GTP exchange protein [imported] - fission yeast (Schizosaccharomyces

C:Species: Schizosaccharomyces pombe

C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000

C:Accession: T50454

R:McDougall, R.C.; Rajadream, M.A.; Barrell, B.G.; Simmonds, M.; Churcher, C.M.

submitted to the EMBL Data Library, November 1999

A:Reference number: Z25030

A:Accession: T50454

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1158 <MCD>

A:Cross-references: EMBL:AL132828; PIDN:CAB60236.1; GSPDB:GN00066; SPDB:SPAC1006.06

A:Experimental source: strain 972h(-); cosmid cl006

C:Genetics:

A:Gene: SPDB:SPAC1006.06

A:Map position: 1

A:Introns: 835/1; 975/3

C:Superfamily: CDC24 homology

Query Match 11.9%; Score 121.5; DB 2; Length 1158;

Best Local Similarity 23.1%; Pred. No. 0.019;

Matches 45; Conservative 36; Mismatches 101; Indels 13; Gaps 3;

QY 4 EVSTTERTYKDLVITSWFQSTVSKEDAMPEALKS----LIFPNFELPKHFTNFKLEI 59

DB 454 EVIYTERDFVRDLEYIRDFWIKPFLSTSNVIPENNROQFIRCVFHNMQIH-----AV 505

QY 60 EORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSSRRLE 119

DB 506 NSRLSNALNRTQLQPVVNTIGLFDYVPKFPFPTKYGANQIAKFEFEREKSTNRNFA 565

QY 120 NFRDPE-LQKVCYLPNTFLRLHRLMHYKQVLERLCKKHPPSHADFRDCRAALAEIT 178

DB 566 NYVHEVERLRESKLENGYLKPTTRLARYPLLSSVLKYTDKNDPDENIPRVEMIR 625

QY 179 EMVAQLHGTMIKMEN 193

DB 626 EFLTILNYETGKTEN 640

RESULT 11

S29083

guanine-nucleotide-releasing protein - rat

N:Alternate names: CDC25 protein homology

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999

C:Accession: S29083

R:Shou, C.; Farnsworth, C.L.; Neel, B.G.; Feig, L.A.

Nature 358, 351-354, 1992

A:Title: Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor fo

A:Reference number: S29083; MUID:92350260

A:Accession: S29083

A:Molecule type: mRNA

A:Residues: 1-1244 <SHO>

A:Cross-references: EMBL:X67241; NID:g57664; PIDN:CAA47666.1; PID:g57665

A:Note: the authors translated the codon GAG for residues 135 and 137 as Gln

C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol

F:240-426/Domain: CDC24 homology <CD24>

F:1005-1241/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 11.7%; Score 120; DB 2; Length 1244;

Best Local Similarity 22.1%; Pred. No. 0.029;

Matches 44; Conservative 33; Mismatches 90; Indels 32; Gaps 5;

QY 9 ERTYKDLVITSWF-----QSTVSKEDAMPEALKSLIFPNFELPKHFTNFKLEIQR 63

DB 252 EAEYVQQLHILVNNFLRPLRMAASSKKPPITHDDVSSIFLNSETIMFLHQIFYQGLKARI 311

QY 64 ALWEGRSNAQIRD-----YORIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSS 115

DB 312 ASWPTLVADLFDILLPMLNIYQ-----FVRNHQYSLQILAH-----CKQN 353

QY 116 RLLENFCRDFELQKVC-YLPLNTFLRLHRLMHYKQVLERLCKKHPPSHADFRDCRAAL 174

DB 354 RQFDKLLKQYKAPDCERTLETFTLTPMFIPIRYITLHLLAHTPHEHVERNSLDYAK 413

QY 175 AEITENVAQLHGTMIKMEN 193

DB 414 SKLELSRYVMHDEVSETEN 432

RESULT 12

T34428

hypothetical protein F55C7.7d - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34428

R:Du, Z.; Le, T.

submitted to the EMBL Data Library, November 1996

A:Description: The sequence of C. elegans cosmid F55C7.

A:Reference number: Z21524

A:Accession: T34428

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-552 <DUZ>

A:Cross-references: EMBL:U80436; PIDN:AC71111.1; GSPDB:GN00019; CESP:F55C7.7d

A:Experimental source: strain Bristol N2; clone F55C7

C:Genetics:

A:Gene: CESP:F55C7.7d

A:Map position: 1

A:Introns: 15/2; 56/3; 329/3; 415/3; 494/2

Query Match 11.6%; Score 119; DB 2; Length 552;

Best Local Similarity 25.7%; Pred. No. 0.013;
Matches 32; Mismatches 62; Indels 30; Gaps 5;

QY 1 IAREVSTERTYLKDLFVITSWFQSTVSYSKDAMPEAL---KSLLFFNFEPFLHKPH-TNFL 56
 : : : : | : | | : : : | : | : : : | : | : : : | : | :
Db 167 VLMVELTQDYVKDLSVVVEGYIGNLKNMD-LPADLVGDKKIIFANIVNILEFHKTNFL 225

QY 57 KETEORIALWEGSRNAOIRDYQRIGDVMLKNIOGMKHAAHLWKHSEALENGIKSSR 116
 ||| : : | : | :
Db 226 KEIEKCSENTEAAGAPVKYERRL-----HTLVITYCQNKKPSDY 265

QY 117 RL-----ENFCRFELQKVCLPLNTFLTLLRHRLMHYKOVLERLCK 158
 | : | : : : : : : : | : | : | : | : | : | : | : | : | : | :
Db 266 LLAQDDFEAFADTKAKGLGHKVACDLLLIKPVOKIMKYQLLKDIKL 312

RESULT 13

T42739

guanine nucleotide exchange factor unc-73a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000

C:Accession: T42739; T42740; T34427; T34430; T34426

R:Steven, R.; Kubiseski, T.J.; Zheng, H.; Kulkarni, S.; Mancillas, J.; Ruiz Morales, A.; Cell 92, 785-795, 1998

A:title: UNC-73 activates the Rac GTPase and is required for cell and growth cone migration submitted to the EMBL Data Library, November 1996

A:Reference number: Z22256; MUID:98188103

A:Accession: T42739

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-2488 <STE>

A:Cross-references: EMBL:AF048834; NID:g2944395; PIDN:AAC12931.1; PID:g2944396

A:Accession: T42740

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1634,'QN','E'<SWZ>

A:Cross-references: EMBL:AF048835; NID:g2944397; PIDN:AAC12932.1; PID:g2944398

R:Du, Z.; Le, T.

A:Description: The sequence of C. elegans cosmid F55C7.

A:Reference number: Z21524

A:Accession: T34427

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2488 <DUZ>

A:Cross-references: EMBL:U80436; PIDN:AAC71108.1; GSPDB:GN00019; CESP:F55C7.7a

A:Experimental source: strain Bristol N2; clone F55C7

A:Accession: T34430

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1634,'QN','E'<DU2>

A:Cross-references: EMBL:U80436; PIDN:AAC71109.1; GSPDB:GN00019; CESP:F55C7.7b

A:Experimental source: strain Bristol N2; clone F55C7

A:Accession: T34426

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 'MVTKCFM',1568,'LRK',1572-2488 <DU3>

A:Cross-references: EMBL:U80436; PIDN:AAC71110.1; GSPDB:GN00019; CESP:F55C7.7c

A:Experimental source: strain Bristol N2; clone F55C7

C:Genetics:

A:Gene: unc-73; F55C7.7a; unc-73; F55C7.7b; CESP:F55C7.7c

A:Map position: 1

A:Introns: 7/3; 64/2; 95/3; 230/3; 265/2; 335/3; 360/3; 488/3; 569/1; 695/1; 1028/2; 1061/3; 2262/2; 2331/1; 2351/3; 2396/1; 2431/1

```

Qy 57 KEIQRIALWEGRNAIQIRDYQTDVWLKKNIQGMKHLAAHLWKHSEALENCIKSSR 116
      ||||:  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 1872 KEIEKCSNTEAAGAAAFVKYERRL-----HTLVYTYCNKPKRSDY 1911
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 117 RL-----ENCFRDELOKVCVLPNTFLRLPHRLMHYKQVLERLCK 158
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1912 LLAQDFEAFADTKAKLGHKVALCDLLIKPQRIKMYQLLLKDLK 1958
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 14
TVHUBD
transforming protein dbl mutant fusion protein - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1989 #sequence_revision 15-Mar-1996 #text_change 14-May-1999
C:Accession: A28051
R:Eva, A.; Vecchio, G.; Rao, C.D.; Tronick, S.R.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 2061-2065, 1988
A:Title: The predicted DBL oncogene product defines a distinct class of transforming
A:Reference number: A28051; MUID:88176886
A:Accession: A28051
A:Molecule type: mRNA
A:Residues: 1-478 <EVA>
A:Cross-references: EMBL:J03639
C:Comment: This sequence is the chimeric product of a translocation mutation.
C:Genetics:
A:Gene: -/MCF2
A:Map position: -/Xq26.3-q27.1
C:Keywords: fusion protein; transforming protein
F:1-50/Region: unidentified locus
F:51-478/Region: transforming protein dbl

Query Match 11.5%; Score 118; DB 4; Length 478;
Best Local Similarity 21.08; Pred. No. 0.013;
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;

Qy 1 IAKEVSTERTYLDLEVTISWFQSTVSKE---DAMPEAL---KSLIFPNFEPLHKFHTN 54
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 52 VLNELIQTERVYVRELYTVLLGYRAEMDNPMFDLMPPLRNKKDILFGNMAIYEFHND 111
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 55 FLKEIQRIALWEGRNAIQIRDYQRIQIGDVMLKNIQGMKHLAAH-----LW-KHSEA 104
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 112 IF-----LSSLENAHAP-----ERVGCFLEKDDDFQYAKYCNKPKRSETIMRKYSEC 161
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 105 LEALENGIKSSRRLENFCRDFELQKVCVLPNTFLRLPHRLMHYKQVLERLCKHHPPSH 164
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 162 AFQECORKLKHRLR-----LDSVLLKPKVQRIKYQLLLKELLKYS-----202
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Qy 165 ADFRDC-----RAALAEITEMVAQLHGTM 188
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 203 ---KDCEGSALLKALDAMLDDLKSVNSDM 229
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 15
TVHUBD
transforming protein dbl precursor - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 02-Sep-1997
C:Accession: A30040
R:Ron, D.; Tronick, S.R.; Aaronson, S.A.; Eva, A.
EMBO J. 7, 2465-2473, 1988
A:Title: Molecular cloning and characterization of the human dbl proto-oncogene: evid
A:Reference number: A30040; MUID:89052660
A:Accession: A30040
A:Molecule type: mRNA
A:Residues: 1-925 <RON>
A:Cross-references: EMBL:X12556
C:Genetics:
A:Gene: GDB:MCF2
A:Cross-references: GDB:120168; OMIM:311030
A:Map position: Xq26.3-Xq27.1
C:Superfamily: dbl transforming protein; CDC24 homology; pleckstrin repeat homology
C:Keywords: oncogene; phosphoprotein; transforming protein

```

F;495-675/Domain: CDC24 homology <CD24>
 F;498-925/Product: transforming protein dbl #status predicted <DBL>
 F;703-807/Domain: pleckstrin repeat homology <PLK>

```

Query Match      11.5%; Score 118; DB 1; Length 925;
Best Local Similarity 21.0%; Pred. No. 0.03;
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;

QY  1 IAKEVSTTERTYLLKDLVITSNFQSTVSKE---DAMPEAL---KSLIFPNFELPHKFTN 54
   : ||| |::| : : : : | || | | : : | : : | :
Db  499 VLNELIQTERVYVRELYTVLLGYRAEMDNPEFDLMPPLRNKKDILFGNMAEYIEFHND 558
   : | : | : | : | : : : | : | : | : | : | :
QY  55 FLKEIEQRLALWEGRSNAQIRDYRIGDVMKNIQGMKHLAAH-----LW-KHSEA 104
   : | : | : | : | : | : | : | : | : | : | :
Db  559 IF-----LSSLENCAP-----ERYGPGFLERKDDFQMYAKYCNKPRSEIWRKYSEC 608
   : | : | : | : | : | : | : | : | : | : | :
QY  105 LEALENGIKSSRRLENFCRDFELQKVCYLPPLNTFLRLPLRLMHYKQVLERLCKKHPPSH 164
   : | : | : | : | : | : | : | : | : | : | :
Db  609 AFFEQCQKILKHLR-----LDSYLLKPVQRITKYQLLLKELIKYS----- 649
   : | : | : | : | : | : | : | : | : | : | :
QY  165 ADPRDC-----RAALAEITEMVAQLHGTM 188
   : || : || : : : : : : :
Db  650 ---KCEGSALLKKALDAMDLLKSYNDSM 676
   : | : | : | : | : | : | : | : | : | : | :
  
```

Search completed: December 6, 2001, 08:52:39
 Job time: 371 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:53:08 ; Search time 28.69 Seconds
(without alignments)
247.925 Million cell updates/sec

Title: US-09-555-342A-2_COPY_544_737

Perfect score: 1023

Sequence: 1 IAEKSTERTYKDLVIT.....AEITEMVAQLHGTMKMF 194

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	ID	Description
1	154	15.1	960	1 FGD1_MOUSE	P52734 mus musculus
2	154	15.1	961	1 FGD1_HUMAN	P98174 homo sapien
3	138	13.5	3038	1 TRIO_HUMAN	O75962 homo sapien
4	131	12.8	646	1 Y142_HUMAN	Q14155 homo sapien
5	129.5	12.7	596	1 LFC_MOUSE	Q60875 mus musculus
6	129.5	12.3	893	1 GEFH_HUMAN	Q92974 homo sapien
7	120	11.7	1244	1 GNRP_RAT	P28818 rattus norv
8	118	11.5	925	1 DBL_HUMAN	P10911 homo sapien
9	117.5	11.5	738	1 ECT2_MOUSE	Q07139 mus musculus
10	117.5	11.5	2044	1 SIF1_DROME	P91620 drosophila
11	117.5	11.5	2064	1 SIF1_DROME	P91621 drosophila
12	116	11.3	872	1 SCD1_SCHPO	P40995 schizosacch
13	116	11.3	1262	1 GNRP_MOUSE	P27671 mus musculus
14	114	11.1	854	1 CC24_YEAST	P11433 saccharomyc
15	111	10.9	937	1 DBS_RAT	Q63406 rattus norv
16	111	10.9	1275	1 GNRP_HUMAN	Q13972 homo sapien
17	109.5	10.7	1108	1 DBS_HUMAN	O15068 homo sapien
18	108	10.6	845	1 VAV_MOUSE	P27870 mus musculus
19	106	10.4	845	1 VAV_HUMAN	P15498 homo sapien
20	105.5	10.3	1693	1 RIP2_MOUSE	P97433 mus musculus
21	105	10.3	1271	1 BCR_HUMAN	P11274 homo sapien
22	104	10.2	1591	1 TIAM_HUMAN	Q13099 homo sapien
23	104	10.2	1591	1 TIAM_MOUSE	Q60610 mus musculus
24	103.5	10.1	847	1 VAV3_HUMAN	Q9UKW4 homo sapien
25	103	10.1	843	1 VAV_RAT	P54100 rattus norv
26	103	10.1	878	1 VAV2_HUMAN	P52735 homo sapien
27	102	10.0	868	1 VAV2_MOUSE	Q60992 mus musculus
28	101	9.9	519	1 TIAM_HUMAN	Q12774 homo sapien
29	99	9.7	1663	1 HAPB_HUMAN	O60229 homo sapien
30	98	9.6	1919	1 HAPB_RAT	P97924 rattus norv
31	96.5	9.4	2137	1 SPCB_HUMAN	P11277 homo sapien
32	95	9.3	1149	1 DBS_MOUSE	Q64096 mus musculus
33	94.5	9.2	847	1 VAV3_MOUSE	Q9P0C8 mus musculus

RESULT 1

ID	FGD1_MOUSE	STANDARD;	PRT;	960 AA.
AC	P52734;			
DC	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)			
DE	(FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).			
GN	FGD1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96081343; PubMed=8535076;			
RA	Pasteris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,			
RA	Gorski J.L.			
RT	"Cloning and regional localization of the mouse facio-genital			
RT	dysplasia (Fgdl) gene.;"			
RL	Mamm. Genome 6:658-661(1995).			
CC	!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS			
CC	BY EXCHANGING BOUND GDP FOR FREE GTP.			
CC	!- SURCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).			
CC	!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).			
CC	!- SIMILARITY: CONTAINS 2 PH DOMAINS.			
CC	!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U22325; AAA96001.1; -			
DR	MGI; MGI:104566; Fgdl.			
DR	InterPro; IPR001849; PH.			
DR	InterPro; IPR000219; RhoGEF.			
DR	InterPro; IPR000822; Znf-C2H2.			
DR	InterPro; IPR000306; Znf_FYVE.			
DR	Pfam; PF01363; FYVE; 1.			
DR	Pfam; PF00169; PH; 2.			
DR	Pfam; PF00621; RhoGEF; 1.			
DR	SMART; SM00064; FYVE; 1.			
DR	SMART; SM00233; PH; 2.			
DR	SMART; SM00325; RhoGEF; 1.			
DR	PROSITE; PS50178; 2F_FYVE; 1.			
DR	PROSITE; PS50003; PH_DOMAIN; 2.			
KW	Guanine-nucleotide releasing factor; zinc-finger.			
FT	DOMAIN 7 330			
FT	DOMAIN 171 179			
FT	DOMAIN 179 187			
FT	DOMAIN 589 688			
FT	SH3-BINDING (POTENTIAL).			
FT	SH3-BINDING (POTENTIAL).			
FT	PH 1.			

ALIGNMENTS

```
FT ZN_FING 729 789 FIVE-TYPE.
SQ DOMAIN 820 920 PH 2.
SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 15.1%; Score 154; DB 1; Length 960;
Best Local Similarity 23.7%; Pred. No. 1.le-05;
Matches 47; Conservative 41; Mismatches 96; Indels 14; Gaps 4;

QY 1 IAEKSTERTYTKDLEVTISWFSQSVSKE----DAMPEALKSLIFPNFEPLHKFTNF- 55
Db 376 IANELLQTEKAYSVRLHLLDQVFCARLLEARNRSPADVVGIFSNICSICYCFHQQFL 435
QY 56 LKEIEQRLALWEGRSNAQIRDYQRIQDVMKNIQGMKHLAAHLWKHSEALEALENGIKSS 115
Db 436 LPELEKRMEDW-----RYPRIGDILQKLAPFLKMYGVKNFDRAVELVNTWTERS 487
QY 116 RLLENFCRDFELQKVC-YLPLNTFLRLPHLMHYQVLERLCKKHPPSHAFDRCAAL 174
Db 488 TQFKVIHEVQKEACGNLTQHHMLPEVQIPRYVELLLKDYLLKLPHGSPDSKDAQSL 547
QY 175 AEITENVAQLHGTMIKME 192
Db 548 ELIATAEHSNAIRKME 565

RESULT 2
FGDL_HUMAN STANDARD; PRT; 961 AA.
ID FGDL_HUMAN STANDARD; PRT; 3038 AA.
AC P98174;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN).
GN FGDL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Craniofacial;
RX MEDLINE=95042764; PubMed=7954831;
RA Pasternis N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,
RA Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.;
RT "Isolation and characterization of the facio-genital dysplasia
RT (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide
RT exchange factor."
RL Cell 79:669-678(1994).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GTP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA,
CC AND LESS IN LIVER; ADULT HEART, BRAIN, LUNG, SKELETAL MUSCLE, AND
CC LESS IN PANCREAS AND LIVER.
CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE
CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT
CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FIVE-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U11690; AAA57004.1;
CC MIM; 305400;
CC
CC DR
CC *DR
```

```
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000822; Znf-C2H2.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 2.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 2.
DR SMART: SM00325; RhoGEF; 1.
DR PROSITE: PS50178; ZF_FYVE; 1.
DR PROSITE: PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 590 689 PH 1.
FT ZN_FING 730 790 FYVE-TYPE.
FT DOMAIN 821 921 PH 2.
SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match 15.1%; Score 154; DB 1; Length 961;
Best Local Similarity 23.7%; Pred. No. 1.le-05;
Matches 47; Conservative 41; Mismatches 96; Indels 14; Gaps 4;

QY 1 IAEKSTERTYTKDLEVTISWFSQSVSKE----DAMPEALKSLIFPNFEPLHKFTNF- 55
Db 377 IANELLQTEKAYSVRLHLLDQVFCARLLEARNRSPADVVGIFSNICSICYCFHQQFL 436
QY 56 LKEIEQRLALWEGRSNAQIRDYQRIQDVMKNIQGMKHLAAHLWKHSEALEALENGIKSS 115
Db 437 LPELEKRMEDW-----RYPRIGDILQKLAPFLKMYGVKNFDRAVELVNTWTERS 488
QY 116 RLLENFCRDFELQKVC-YLPLNTFLRLPHLMHYQVLERLCKKHPPSHAFDRCAAL 174
Db 489 TQFKVIHEVQKEACGNLTQHHMLPEVQIPRYVELLLKDYLLKLPHGSPDSKDAQSL 548
QY 175 AEITENVAQLHGTMIKME 192
Db 549 ELIATAEHSNAIRKME 566

RESULT 3
TRIO_HUMAN STANDARD; PRT; 3038 AA.
ID TRIO_HUMAN STANDARD; PRT; 3038 AA.
AC Q75962; Q13458;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRIPLE FUNCTIONAL DOMAIN PROTEIN (PTPRF INTERACTING PROTEIN).
GN TRIO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Fibroblast;
RX MEDLINE=96224308; PubMed=8643598;
RA Debat A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA Streuli M.;
RT "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT phosphatase, contains a protein kinase domain, and has separate rac-
RT specific and rho-specific guanine nucleotide exchange factor
RT domains."
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
[2]
RN RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Streuli M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
```



```
Db 2012 IVSEYIDTFPDLKORLGHRLQTLTDLLIKPVQRMKYQLLKLKFLKYSKASLDTSER 2071
QY 173 AL 174
   | :
Db 2072 AV 2073

RESULT 4
Y142_HUMAN
ID Y142_HUMAN STANDARD; PRT; 546 AA.
AC Q14155;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0142.
GN KIAA0142.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D63476; BAA09763.1; -.
DR HSSP; P19174; IHSQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
KW Hypothetical protein, SH3 domain.
FT DOMAIN 6 65
FT DOMAIN 295 400
FT SEQUENCE 646 AA; 73140 MW; B5B5A83F0EBC28D2 CRC64;

Query Match 12.8%; Score 131; DB 1; Length 646;
Best Local Similarity 25.1%; Pred. No. 0.00064;
Matches 49; Conservative 31; Mismatches 81; Indels 34; Gaps 5;

QY 1 IAKEVSTERTYTKDLKLEVTISWFSQSVSKEDAMPEALKSLIPFNPLKFTNFKETE 60
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 VLQNILETENEYSKELOTVLSYLRPLQTSKISSANISYLMGNLEICSFOQMLVQSLE 156
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 QRUALMEGRSNAQIRDYQRIGDVMLKNIQGMKHL-----AAH-----LWKHSEAL-EAL 108
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 ECTKLPEAQ-----QRVGGCFLMLPQMKTLTYLTICANHPSAVNLTSEELGFM 208
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 109 ENGIKSSRLNEFCRDFELQKVCYLPLNTFLRLPLRLMHYKQVLRCLCKHHPPSHADFR 168
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 209 ETKGASSPGI-----LVLTTCLSKPFMRDLKDYPTLLKELERHMDYHTDRQ 254
QY 169 DCRAALAEITEMVAQ 183
   | : | |
Db 255 DIQKSMAAFKNLSAQ 269

RESULT 5
LFC_MOUSE
ID LFC_MOUSE STANDARD; PRT; 596 AA.
AC Q60875; O09115;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LYMPHOID BLAST CRISIS-LIKE 1 (LBC'S FIRST COUSIN) (ONCOGENE LFC)
DE (RHOBIN).
GN LBC1L OR LFC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid;
RX MEDLINE=95355462; PubMed=7629163;
RA Whitehead I., Kirk H., Tognon C., Trigo-Gonzalez G., Kay R.;
RT "Expression cloning of lfc, a novel oncogene with structural
RT similarities to guanine nucleotide exchange factors and to the
RT regulatory region of protein kinase C";
RL J. Biol. Chem. 270:18388-18395(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Olofsson B.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS, WITH THE EXCEPTION OF LIVER
CC TISSUE. LEVELS ARE HIGH IN HEMOPOIETIC TISSUES (THYMUS, SPLEEN,
CC BONE MARROW) AS WELL AS IN KIDNEY AND LUNG.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -!- SIMILARITY: TO HUMAN NUCLEOTIDE EXCHANGE PROTEIN LBC.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U28495; AAC52234.1; -.
DR EMBL; X95761; CAA65067.1; -.
DR MGD; MGI:103284; Lbcll.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene;
KW Phorbol-ester binding; Zinc.
FT DOMAIN 40 86
FT DOMAIN 240 511
FT DOMAIN 487 573
FT DOMAIN 493 496
FT POLY-LEU.
```

```
FT CONFLICT 1 21 MSRIEQLTRARIDRSKEQATK -> MSGNRRPSPRRGQ
FT CONFLICT 156 156 A -> V (IN REF. 2).
FT CONFLICT 574 596 MISSING (IN REF. 1).
SQ SEQUENCE 596 AA; 68584 MW; E168508BDC7C6E13 CRC64;

Query Match 12.7%; Score 129.5; DB 1; Length 596;
Best Local Similarity 21.4%; Pred. No. 0.00078;
Matches 44; Conservative 50; Mismatches 85; Indels 27; Gaps 7;

QY 4 EVSTTERTYKLEVITSNFQSVSKEDAM-PEALKSLIFPNEPLKHFHTNFKEL--E 60
Db 243 ELTQTELHHVTRLKIMTRLFRTGMLEELHLEPGVVOGL-FPCVDELSDTHTRFLNQLLR 301

QY 61 QRLAWEGRSNAQIRDYQIRIGDVMKNIQ-----MKHLAAHLWKHSEALENGIKSS 114
Db 302 RQALCPGSTRNFV--IHLRGDLLISQFSGSNAEQMRKTYSEFCRSHTALKLYKELYAR 359

QY 115 SRLENFCRDFELQKV-----CYLPLNTLLRPLRLMHYKQVLERCKLKHPPSHAD 166
Db 360 DRFOQFIRKTRSAVILKRHGQEC-----ILLVTRITKYPVLINILQNSHGVEE 412

QY 167 FRDCRAALAEITEMVAQLHGTWKME 192
Db 413 YQDLASALGLVKELLSNVDDVHELE 438

RESULT 6
GEPH_HUMAN
ID GEPH_HUMAN STANDARD; PRT; 893 AA.
AC Q2974; Q15079; 075142;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GEF-H1 PROTEIN (PROLIFERATING CELL NUCLEOLAR ANTIGEN P40).
GN LFP40 OR KIA0651.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99074271; PubMed=9857026;
RA Ren Y., Li R., Zheng Y., Busch H.;
RT "Cloning and characterization of GEF-H1, a microtubule-associated
RT guanine nucleotide exchange factor for Rac and Rho GTPases.";
RL J. Biol. Chem. 273:34954-34960(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [3]
RP REVISIONS.
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 630-893 FROM N.A.
RX MEDLINE=89168219; PubMed=2466560;
RA Reddy A.B., Chatterjee A., Rothblum L.I., Black A., Busch H.;
RT "Isolation and characterization of complementary DNA to proliferating
RT cell nucleolar antigen P40.";
RL Cancer Res. 49:1763-1767(1989).
CC -!- FUNCTION: STIMULATES GUANINE NUCLEOTIDE EXCHANGE OF RAC AND RHO
CC BUT IS INACTIVE TOWARD CDC42, TC10, OR RAS. BINDS TO RAC AND RHO
CC PROTEINS IN BOTH THE GDP- AND GUANOSINE 5'-3'-O-(THIO)TRIPHOSPHATE-
```

```
CC BOUND STATES WITHOUT DETECTABLE AFFINITY FOR CDC42 OR RAS. MAY
CC HAVE A DIRECT ROLE IN ACTIVATION OF RAC AND/OR RHO AND IN BRINGING
CC THE ACTIVATED GTPASE TO SPECIFIC TARGET SITES SUCH AS
CC MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: COLOCALIZES WITH MICROTUBULES THROUGH THE
CC CARBOXYL-TERMINAL COILED-COIL DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -!- SIMILARITY: COULD BE THE ORTHOLOG OF MOUSE LYMPHOID BLAST CRISIS-
CC LIKE 1 (LFC ONCOGENE). SHOWN HERE COMES FROM THE FIGURE OF REF.1,
CC CAUTION: THE SEQUENCE SUBMITTED HERE IS WRONG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U72206; AAC97383.1; ALT_SEQ.
DR EMBL; AB014551; BAA31626.2; -.
DR EMBL; X15610; CAA33634.1; -.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Phorbol-ester binding; Zinc;
KW Coiled coil.
FT DOMAIN 40 86 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 238 509 DH.
FT DOMAIN 485 571 PH.
FT DOMAIN 589 610 COILED COIL (POTENTIAL).
FT DOMAIN 797 868 COILED COIL (POTENTIAL).
FT CONFLICT 1 21 MSRIEQLTRARIDRSKEQATK -> LGSAHAGMCCCGCCP
FT CONFLICT 193 193 LLAHLEOGSLRR (IN REF. 2).
FT CONFLICT 579 579 E -> EA (IN REF. 2).
FT CONFLICT 867 867 S -> P (IN REF. 2).
FT CONFLICT 885 885 Q -> P (IN REF. 2).
FT CONFLICT SSPQAMPCT -> SLPAGDALYLSFNPPQPSRGTDRLDLPV
FT SEQUENCE 893 AA; 101173 MW; 80AC8FA7F762E9C3 CRC64;
SQ SEQUENCE 893 AA; 101173 MW; 80AC8FA7F762E9C3 CRC64;

Query Match 12.3%; Score 125.5; DB 1; Length 893;
Best Local Similarity 20.5%; Pred. No. 0.0027;
Matches 41; Conservative 51; Mismatches 93; Indels 15; Gaps 5;

QY 4 EVSTTERTYKLEVITSNFQSVSKEDAMPEALKSLIFPNEPLKHFHTNFKEL--EQ 61
Db 241 ELTQTELHHVTRLKIMTRLFRTGMLEELHLEPGVVOGLFPCVDELSDTHTRFLNQLLR 300

QY 62 RIALWEGRSNAQIRDYQIRIGDVMKNIQ-----MKHLAAHLWKHSEALENGIKSS 115
Db 301 RQALCPGSTRNFV--IHLRGDLLISQFSGSNAEQMRKTYSEFCRSHTALKLYKELYAR 358

QY 116 RRLNFCRDFELQKVCLP---INTFLRPLRLMHYKQVLERCKLKHPPSHADFRDCA 172
Db 359 KRQEQIR--KYTRPAVLKRHGQECILLVTRITKYPVLINILQNSHGIEERDILT 416

QY 173 ALAEITEMVAQLHGTWKME 192
Db 413 YQDLASALGLVKELLSNVDDVHELE 438
```


J. Biol. Chem. 269:62-65(1994).
-!- FUNCTION: PROMOTES THE EXCHANGE OF G25K-BOUND GDP BY GTP. ACTS AS
A GUANINE NUCLEOTIDE EXCHANGE FACTOR FOR CDC42 AND RHOA, BUT NOT
FOR RAC1 AND TC10. CAN BE ACTIVATED BY TRUNCATION OF THE N-
TERMINUS.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
-!- TISSUE SPECIFICITY: EXPRESSED IN FETAL BRAIN AND ADRENAL GLANDS;
IN THE ADULT, EXPRESSION IS LIMITED TO TESTES, OVARIES AND BRAIN.
-!- DOMAIN: THE DH DOMAIN IS ESSENTIAL FOR TRANSFORMING ACTIVITY AND
DIRECTLY CATALYZES GDP-GTP EXCHANGE ACTIVITY.
-!- DISEASE: MCF.2 AND DBL REPRESENT TWO ACTIVATED VERSIONS OF THE
SAME PROTO-ONCOGENE.
-!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
-!- SIMILARITY: CONTAINS 1 PH DOMAIN.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

DR EMBL; X12556; CAA31069.1; -
DR EMBL; J03639; AAS2172.1; ALT_INIT.
DR EMBL; X13230; CAA31617.1; ALT_SEQ.
DR PIR; A30040; TVHUBD.
DR PIR; A28051; TVHUBD.
DR MIM; 311030; -
DR InterPro; IPR000947; CBFA_NFYB.
DR InterPro; IPR001251; CRAL_TRIO.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00150; SPEC; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene; Phosphorylation;
Transforming protein.
FT CHAIN 398 925 MCF2 TRANSFORMING PROTEIN.
FT CHAIN 498 925 DBL TRANSFORMING PROTEIN.
FT DOMAIN 498 686 DH.
FT DOMAIN 687 809 PH.
FT MOD_RES 742 742 PHOSPHORYLATION.
FT MUTAGEN 640 646 CAPABILITY REDUCED; NO STIMULATION OF GDP
DISSOCIATION.
FT CONFLICT 634 634 R -> Q (IN REF. 4).
FT CONFLICT 886 886 A -> V (IN REF. 3).
FT SEQUENCE 925 AA; 107657 MW; 0BCCB303276D19C9 CRC64;

Query Match 11.5%; Score 118; DB 1; Length 925;
Best Local Similarity 21.0%; Pred. No. 0.013;
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;

QY 1 IAKEVSTERTYKLDLEVTISWFQSTVSKE---DAMPEAL---KSLFPNPEPLKHFHTN 54
Db 499 VLNELIQTRVYVRELYTVLLGYSRAEDNPEDFLPPLRNKKDKILFGNMAEYEFHND 558
QY 55 FLKEIQRIALWEGRNAQIRFQTDVWLKNIOGKKHLAAH-----LW-KHSEA 104
Db 559 IF-----LSSLENCAHAP-----ERVGPCFLERKDDFQMYAKYCKQNKPRSETIWRKYSEC 608
QY 105 LEALENGIKSSRLENCFRDFELQKVCYLPPLNTFLRLPLRLMHYKQVLERLCKKHPPSH 164
Db 609 AFFQECQRLKHLR-----LDSYLLKPQVIRITYQLLLKELLKYS----- 649

QY 165 ADRPDC-----RAALAEITEMVAQLHGTM 188
Db 650 ---KDCEGSALLKALDAMDLLKSVNDMSM 676

RESULT 9
ECT2_MOUSE STANDARD; PRT; 738 AA.
ID ECT2_MOUSE AC Q07139;
DT 15-JUL-1999 (rel. 38, Created)
DT 15-JUL-1999 (rel. 38, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE ECT2 PROTEIN (ECT2 ONCOGENE).
GN ECT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93218723; PubMed=8464478;
RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;
RT "Oncogene ect2 is related to regulators of small GTP-binding
RT proteins.";
RL Nature 362:462-465(1993).
RN [2]
RP ERRATUM.
RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;
RL Nature 364:737-737(1993).
CC -!- FUNCTION: TRANSFORMING PROTEIN THAT BINDS HIGHLY SPECIFICALLY TO
CC RHOA, RHOA AND RAC PROTEINS, BUT DOES NOT APPEAR TO CATALYZE
CC GUANINE NUCLEOTIDE EXCHANGE. TRUNCATION OF THE N-TERMINUS PUTS THE
CC PROTEIN IN AN ACTIVATED STATE.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS. ALSO DETECTABLE
CC IN KIDNEY, LIVER AND SPLEEN.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

DR EMBL; L11316; AAA37536.1; -
DR MGD; MGI:95281; Ect2.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00533; BRCT; 2.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00292; BRCT; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS0172; BRCT; 2.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene.
FT DOMAIN 280 499 DH.
FT DOMAIN 500 619 PH.
FT SEQUENCE 738 AA; 83685 MW; 3D270CBF1570C681 CRC64;

Query Match 11.5%; Score 117.5; DB 1; Length 738;
Best Local Similarity 22.3%; Pred. No. 0.011;
Matches 50; Conservative 30; Mismatches 79; Indels 65; Gaps 8;

QY 1 IAKEVSTERTYKLDLEVTISWFQSTVSKE-----AMPEALKSLFPNPEPLKHFHTN 54
Db 499 VLNELIQTRVYVRELYTVLLGYSRAEDNPEDFLPPLRNKKDKILFGNMAEYEFHND 558

Db 281 VAKELYTESNYVILATIIQLFOVPLEEGRGPGILAPEIKT-IFGSIPIDFVHMK 339
QY 55 FLKEIEQRLALW-EGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALELENGIK 113
Db 340 IKDDLELIANWDESR-----IGDIFLYKAKDL-----VK 370
QY 114 SSRLENFCRDFELQVKY-----CYLPLN-----TFLLRPLRLML 148
Db 371 TYPEPVNF---FEMSEKMIKCEKQKPRFHAFLKINAKPECGRSQSLVELLIRPVQLRPS 427
QY 149 YKQVLERLCKHPPSHADFRCRAALAEITEMVAQLHGWTWIKME 192
Db 428 VALLNDLKKHTADENPKSTLEKAIGSLKEVMTNIEDRRKTE 471

RESULT 10
SIF2_DROME STANDARD; PRT; 2044 AA.
AC P91620;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE STILL LIFE PROTEIN TYPE 2 (SIF TYPE 2).
GN SIF.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RC TISSUE=Head;
RP ERRATUM.
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,
RA Saigo K., Nabeshima Y.-I., Hama C.;
RL Science 275:1405-1405(1997).
CC -1- FUNCTION: REGULATES SYNAPTIC DIFFERENTIATION THROUGH THE
CC ORGANIZATION OF ACTIN CYTOSKELETON POSSIBLY BY ACTIVATING RHO-LIKE
CC GTPASES. IS LIKELY A FACTOR IN THE CASCADE OF RAC1 OR CDC42 IN THE
CC NEURONS.
CC -1- SUBCELLULAR LOCATION: LOCALIZES TO THE SUBMEMBRANOUS REGION OF
CC SYNAPTIC TERMINALS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIF TYPE 1 (AC P91621) AND SIF
CC TYPE 2 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: AT STAGE 14, EXPRESSION OCCURS IN EACH
CC SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION
CC BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND
CC VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.
CC -1- SIMILARITY: CONTAINS 1 DEL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86546; BAA13108.1; -
CC FlyBase; FBgn0019652; sif.
CC InterPro; IPR001131; GDS_CDC24.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001849; PH.
CC InterPro; IPR003116; RBD.

DR InterPro; IPR000219; RhogEF.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF02196; RBD; 1.
DR Pfam; PF00621; RhogEF; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00455; RBD; 1.
DR SMART; SM00325; RhogEF; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Developmental protein; Synapse;
KW Repeat; Alternative splicing. 4 X 25 AA APPROXIMATE REPEAT.
FT DOMAIN 62 249
FT REPEAT 62 86
FT REPEAT 94 118
FT REPEAT 154 178
FT REPEAT 225 249
FT DOMAIN 819 937 PH 1.
FT DOMAIN 1184 1273 PDZ.
FT DOMAIN 1410 1673 DH.
FT DOMAIN 1674 1767 PH 2.
FT DOMAIN 467 470 POLY-PRO.
FT DOMAIN 646 649 POLY-ARG.
FT DOMAIN 1295 1298 POLY-PRO.
FT DOMAIN 1898 1909 POLY-GLN.
FT DOMAIN 1929 1933 POLY-PRO.
SQ SEQUENCE 2044 AA; 228324 MW; 75D7CF21F49654B6 CRC64;

Query Match 11.5%; Score 117.5; DB 1; Length 2044;
Best Local Similarity 21.2%; Pred. No. 0.035;
Matches 41; Conservative 42; Mismatches 99; Indels 11; Gaps 5;

QY 1 IAEKVSTERTYKLDLEIVTSWFQSTVSKEDAMPEALKSLIFNPEPLHKHFNFLKEIE 60
Db 1412 VMELVDTERTYVYKHLNLLLEHYLEPDKMRETFLSNAEINALFGNIHEIVTFQRFQNLQ 1471
QY 61 QRLALMEGRSN-----AQIRDYQ-RIGDVMKNIQGMKHLAAHLWKHSEALELENGIKS 114
Db 1472 ESDDL-EPDENKFEHCQGFNVLFAIGSAFLYVNHFKLYSSFCASHSKAKQVLPN-EG 1529
QY 115 SRLENFCRDFELQVKCYLPLNFTLLRPLRLMLHYKQVLER---LCKHHPSSHADFRCR 171
Db 1530 NHALQEFLEARNPKQSHSSTLESYLKPIQRIKYLKPLLLQOMRNLTDTRADEHVHLCAL 1589
QY 172 AALAEITEMVAQL 184
Db 1590 KGMKVAEHINEM 1602

RESULT 11
SIF1_DROME STANDARD; PRT; 2064 AA.
AC P91621;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE STILL LIFE PROTEIN TYPE 1 (SIF TYPE 1).
GN SIF.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RC TISSUE=Head;
RP SEQUENCE FROM N.A.
RX MEDLINE=97153054; PubMed=89599801;
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,
RA Saigo K., Nabeshima Y.-I., Hama C.;
RT "Still life, a protein in synaptic terminals of Drosophila homologous
RT to GTP-GTP exchangers.";
RL Science 275:543-547(1997).

RT to GDP-GTP exchangers.";
RL Science 275:543-547(1997).
RN [2]
RP ERRATUM.
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,
RA Saigo K., Nabeshima Y.-I., Hama C.;
RL Science 275:1405-1405(1997).
CC -!- FUNCTION: REGULATES SYNAPTIC DIFFERENTIATION THROUGH THE
CC ORGANIZATION OF ACTIN CYTOSKELETON POSSIBLY BY ACTIVATING RHO-LIKE
CC GTPASES. IS LIKELY A FACTOR IN THE CASCADE OF RAC1 OR CDC42 IN THE
CC NEURONS.
CC -!- SUBCELLULAR LOCATION: LOCALIZES TO THE SUBMEMBRANOUS REGION OF
CC SYNAPTIC TERMINALS.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIF TYPE 1 (SHOWN HERE) AND SIF
CC TYPE 2 (AC P91620); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DEVELOPMENTAL STAGE: AT STAGE 14, EXPRESSION OCCURS IN EACH
CC SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION
CC BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND
CC VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86547; BAAL3109.1; -.
CC FlyBase; FBgn0019652; sif.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001849; PH.
CC InterPro; IPR003116; RBD.
CC InterPro; IPR000219; RHOGEF.
CC InterPro; IPR001960; WH1.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF02196; RBD; 1.
CC Pfam; PF00621; RHOGEF; 1.
CC SMART; SM00228; PDZ; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00455; RBD; 1.
CC SMART; SM00325; RHOGEF; 1.
CC SMART; SM00461; WH1; 1.
CC PROSITE; PS00741; GDS_CDC24; 1.
CC PROSITE; PS0106; PDZ; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC Guanine-nucleotide releasing factor; Developmental protein;
CC Myristate; Synapse; Alternative splicing.
CC FT LIPID 2 2 MYRISTATE (POTENTIAL).
CC FT DOMAIN 839 957 PH 1.
CC FT DOMAIN 1204 1293 PDZ.
CC FT DOMAIN 1430 1693 DH.
CC FT DOMAIN 1694 1787 PH 2.
CC FT DOMAIN 445 453 POLY-GLY.
CC FT DOMAIN 545 548 POLY-GLN.
CC FT DOMAIN 1315 1318 POLY-PRO.
CC FT DOMAIN 1918 1929 POLY-GLN.
CC FT DOMAIN 1949 1953 POLY-PRO.
CC SEQUENCE 2064 AA; 230489 MW; D3BDCC10A94D9DE6C CRC64;
Query Match 11.5%; Score 117.5; DB 1; Length 2064;
Best Local Similarity 21.2%; Pred. No. 0.035;
Matches 41; Conservative 42; Mismatches 99; Indels 11; Gaps 5;
QY 1 IAKEVSTTERTYKLDLEVTISFQSTVSKEDAMPEALKSLIFPNFPLKHFHTNFKLEIE 60
DB 1432 VMVELVDVTERTYVKNLNNLEHYLEPKMKRTFLSNALFNINHEIVTFQRFQNLNE 1491

QY 61 QRLALWEGRSN-----AQIRDYO-RIGDVMKNIQGNKHLAHLWKHSEALENGIKS 114
DB 1492 ESULD-EPDNFKPGCGQFRNVLFAIGSAFLYYVNHFKLYSFCASHKSAQKVLHPN-EG 1549
QY 115 SRLENFCRDFELQKVCYPLNTFLRLPLRLHLMHYKQVLER---LCKHHPPSHADFRDCR 171
DB 1550 NHALQEFLLAARNPKQHSSTLESILKPIQRIKLKPLLLQOQMRNLDTTRADEHVHLCAL 1609
QY 172 AALAEITEMVAQL 184
DB 1610 KGMKVAEHINEM 1622
RESULT 12
ID SCD1_SCHPO STANDARD; PRT; 872 AA.
AC P40995;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROTEIN SCD1.
GN SCD1 OR RAL1 OR SPAC16E8.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP870;
RX MEDLINE=95007774; PubMed=7923372;
RA Chang E.C., Barr M., Wang Y., Jung V., Xu H.-P., Wigler M.H.;
RT "Cooperative interaction of S. pombe proteins required for mating and
RT morphogenesis";
RL Cell 79:131-141(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH MOE1, AND SUBCELLULAR LOCATION.
RX MEDLINE=99110922; PubMed=9892665;
RA Chen C.-R., Li Y.-C., Chen J., Hou M.-C., Papadaki P., Chang E.C.;
RT "Moel, a conserved protein in Schizosaccharomycetes pombe, interacts
RT with a Ras effector, Scd1, to affect proper spindle formation.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:517-522(1999).
CC -!- FUNCTION: REQUIRED FOR MATING AND MORPHOGENESIS. MAY CONTAIN A
CC CRYPTIC BINDING SITE FOR CDC42 THAT IS ENHANCED BY BINDING RAS.
CC INTERACTS DIRECTLY WITH SCD2. MAY PROMOTE THE EXCHANGE OF CDC42-
CC BOUND GDP BY GTP (BY SIMILARITY).
CC -!- SUBUNIT: SCD1, SCD2, CDC42, AND RAS1, IN ITS GTP-BOUND STATE, ACT
CC COOPERATIVELY TO FORM A PROTEIN COMPLEX. INTERACTS WITH MOE1.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12538; AAA50556.2; -.
CC EMBL; Z98529; CAB11037.1; -.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR000270; OPR.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RHOGEF.
CC Pfam; PF00564; OPR; 1.

```
DR Pfam; PF00169; PH; 1.
DR SMART; SM00016; RhoGEF; 1.
DR SMART; SM00016; OPR; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Nuclear protein.
FT DOMAIN 426 547 PH
FT DOMAIN 559 667 SER/THR-RICH.
SQ SEQUENCE 872 AA; 99101 MW; 726FAE33B519D69D CRC64;

Query Match
Best Local Similarity 11.3%; Score 116; DB 1; Length 872;
Matches 46; Conservative 35; Mismatches 79; Indels 50; Gaps 6;

QY 1 IAKEVSTTERTYKLDLEIVTSFQSTVSKEDAMPEALKSLIPNPEPLHKHFTNFKETE 60
DB 232 VTAELVETELKYIODELYLSNYMVILOQKILSQDFTLS-IFTNLNILDQRRFLVGL 290
QY 61 QRLALMEGRSNAQIRDYQRLGDVLMKNIQGMKHLAAHLWKHSEALENGIKSSRLEN 120
DB 291 MNLSL-----PVEQRIGLALFI-----ALEEGFSV---YQV 318
QY 121 FCDF-----ELQKVCYL-----PLNTFLRLPLRLMHYKQVLERLCKKHPPS 163
DB 319 FCNFFNAQQLIIDNQNLKLVANLLEPSYELPALLIKPIQRIKCYPLLLNQLLKTPSG 378
QY 164 HADFRCDRAALAEITEMVAQLHGTMIKMEN 193
DB 379 YQVEELKQGMACVVRVANQVNETRIHEN 408

RESULT 13
GNRP_MOUSE
ID GNRP_MOUSE STANDARD; PRT; 1262 AA.
AC P27671.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE EXCHANGE FACTOR CDC25) (CDC25MM).
GN RASGRF1 OR CDC25 OR GRF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RX MEDLINE=93010996; PubMed=1396590;
RA Cen H., Lowy D.B.;
RT "Isolation of multiple mouse cDNAs with coding homology to
RT Saccharomyces cerevisiae CDC25: identification of a region related to
RT Bcr, Vav, Dbl and CDC24."
RL EMBO J. 11:4007-4015(1992).
RN [2]
RP SEQUENCE OF 791-1262 FROM N.A.
RC STRAIN-SWISS; TISSUE=Brain;
RX MEDLINE=92289680; PubMed=1376246;
RA Martegani E., Vanoni M., Zippel R., Coccetti P., Brambilla R.,
RA Ferrarini C., Sturani E.P., Albergina L.;
RT "Cloning by functional complementation of a mouse cDNA encoding a
RT homologue of CDC25, a Saccharomyces cerevisiae RAS activator."
RL EMBO J. 11:2151-2157(1992).
RN [3]
RP SEQUENCE OF 1031-1226 FROM N.A.
RX MEDLINE=92357779; PubMed=1379731;
RA Wei W., Mosteller R.D., Saaval P., Gonzales E., McKinney D.,
RA Dasgupta C., Li P., Liu B.X., Broek D.;
RT "Identification of a mammalian gene structurally and functionally
RT related to the CDC25 gene of Saccharomyces cerevisiae.";
```

```
RL Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20899; AAA02741.1; -.
DR EMBL; X59868; CAA42525.1; -.
DR PIR; S20730; S20730.
DR PIR; S22693; S22693.
DR MGD; MGI:95694; Rasgrf1.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGRE_CDC25.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS50096; IQ; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 22 130
FT DOMAIN 208 233 IQ.
FT DOMAIN 248 459 DH.
FT DOMAIN 460 588 PH 2.
FT DOMAIN 1025 1259 RASGEF.
FT CONFLICT 1033 1033 E -> D (IN REF. 3).
SQ SEQUENCE 1262 AA; 144101 MW; 38BFE68F7C228DC8 CRC64;

Query Match
Best Local Similarity 11.3%; Score 116; DB 1; Length 1262;
Matches 43; Conservative 34; Mismatches 90; Indels 32; Gaps 5;

QY 9 ERTYKLDLEIVTSWF-----QSTVSKEDAMPEALKSLIPNPEPLHKHFTNFKETEORL 63
DB 256 EAEYVOQLHLVNNFLRPLRMASSKKPPITHDDVSSIFLNSETIMFLHQIFQGLKARI 315
QY 64 ALWEGRSNAQIRD-----YORIGDVLMKNIQGMKHLAAHLWKHSEALENGIKSS 115
DB 316 SSWPTLVLAADFLDILLPLMLNIYQ-----FVRNHQYSLQILAH-----CKQN 357
QY 116 RLNEFCRDFELQKVC-YLPLNTFLRLPLRLMHYKQVLERLCKKHPPSHADFRCDRAAL 174
DB 358 RQDFKLLKQYKAPDCERTLETFTYPMFQIPRYILTLHELLAHTPHEHVERNSLDYAK 417
QY 175 AEITEMVAQLHGTMIKMEN 193
DB 418 SKLELSRIMHDEVSETEN 436

RESULT 14
```



```
CC24_YEAST
ID CC24_YEAST STANDARD; PRT; 854 AA.
AC P11433;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CELL DIVISION CONTROL PROTEIN 24 (CALCIUM REGULATORY PROTEIN).
GN CDC24 OR CLS4 OR YAL041W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87277425; PubMed=3301539;
RA Miyamoto S., Ohya Y., Ohsumi Y., Anraku Y.;
RT "Nucleotide sequence of the CLS4 (CDC24) gene of Saccharomyces
RT cerevisiae.";
RL Gene 54:125-132(1987).
RN [2]
RP REVISIONS, AND SIMILARITY TO CDC24 FAMILY.
RX MEDLINE=92095962; PubMed=1755844;
RA Miyamoto S., Ohya Y., Sano Y., Sakaguchi S., Iida H., Anraku Y.;
RT "A DBL-homologous region of the yeast CLS4/CDC24 gene product is
RT important for Ca(2+)-modulated bud assembly.";
RL Biochem. Biophys. Res. Commun. 181:604-610(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF CDC42-BOUND GDP BY GTP.
CC CONTROLS THE POLARITY OF CALMODULIN, AND THE CALCIUM REGULATORY
CC PROCESS OF BUD EMERGENCE. CDC24 MAY BE INVOLVED IN THE INITIAL
CC SELECTION AND ORGANIZATION OF THE BUDDING SITE.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16809; AAA82871.1; -
CC EMBL; U12980; AAC04990.1; -
CC PIR; A27477; A27477.
CC SGD; S0000039; CDC24.
CC InterPro; IPR001715; Calponin_hom.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR000270; OPR.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
CC Pfam; PF00364; OPR; 1.
CC SMART; PF00621; RhoGEF; 1.
CC SMART; SM00033; CH; 1.
CC SMART; SM00016; OPR; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00325; RhoGEF; 1.
CC PROSITE; PS00741; GDS_CDC24; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 281 477 DH.
FT DOMAIN 478 668 PH.
FT DOMAIN 494 600 SER/THR-RICH.
FT DOMAIN 681 778 SER/THR-RICH.
```

```
SQ SEQUENCE 854 AA; 96939 MW; 856BCAF4EB0A67D1 CRC64;

Query Match 11.1%; Score 114; DB 1; Length 854;
Best Local Similarity 22.2%; Pred. No. 0.025;
Matches 49; Conservative 39; Mismatches 81; Indels 52; Gaps 9;

QY 1 IAKEVSTTERTYIKLEVITSWFQSTVSKEDAMPEALKSLIPNFPFLKHFHTNFKLEIE 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 282 IIFEVATERKXVHDLILDKVRQQLDNLITSEEL-YMLFNLGDAIDFQRRFLISLE 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 QRLALWEGRSNAQIR-DYQIRIGDVMKLNIGQMKHLAHLWK-----HSEALEALENG 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 341 -----INALVEPSKQIRGALFMH-----SKHFFKLYEPWSIGNAAIEFLSST 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 112 IKSSR-RLENFCRDFELQKVCLPLNTLLRLPHRLMHYKQVLERLCKHHPHSHADFRD 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 384 LHKMRVDSQSFIIINKLE-----LQSFYKPVQVLCRYPLLVKELLAESSDDN-NTKE 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 CRAAL-----AEITENVAQLHGTMTKMFNF 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 437 LEAALDISKNIARSINENORRTENHOVKKLYGRVYVNWKG 477

RESULT 15
DBS_RAT ID DBS_RAT STANDARD; PRT; 937 AA.
AC Q63406;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR DBS (DBL'S BIG SISTER) (MCF2
DE TRANSFORMING SEQUENCE-LIKE PROTEIN) (OST ONCOGENE) (FRAGMENT).
DE MCF2L OR OST.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Osteosarcoma;
RX MEDLINE=95045368; PubMed=7957046;
RA Horii Y., Beeler J.F., Sakaguchi K., Tachibana M., Miki T.;
RT "A novel oncogene, ost, encodes a guanine nucleotide exchange factor
RT that potentially links Rho and Rac signaling pathways.";
RL EMBO J. 13:4776-4786(1994).
CC -!- FUNCTION: GUANINE NUCLEOTIDE EXCHANGE FACTOR THAT POTENTIALLY
CC LINKS PATHWAYS THAT SIGNAL THROUGH RAC1, RHOA AND CDC42. CATALYZES
CC GUANINE NUCLEOTIDE EXCHANGE ON RHOA AND CDC42 AND INTERACTS
CC SPECIFICALLY WITH THE GTP-BOUND FORM OF RAC1, SUGGESTING THAT IT
CC FUNCTIONS AS AN EFFECTOR OF RAC1. MAY ALSO PARTICIPATE IN AXONAL
CC TRANSPORT IN THE BRAIN. BECOMES ACTIVATED AND HIGHLY TUMORIGENIC
CC BY TRUNCATION OF THE N-TERMINUS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WHERE IT IS
CC FOUND IN NEURONS AND ALPHA-TANCYTES. LOWER LEVEL IN THE HEART.
CC -!- PTM: MAINLY PHOSPHORYLATED ON SERINE.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z35654; CAA84713.1; ALT_INIT.
CC InterPro; IPR001251; CRAL_TRIO.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
```

```

DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00621; RhogEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhogEF; 1.
DR SMART; SM00516; SEC14; 1.
DR SMART; SM00150; SPEC; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene; Phosphorylation.
FT NON_TER 1
FT DOMAIN 499 502 POLY-GLU.
FT DOMAIN 514 692 DH.
FT DOMAIN 716 825 PH.
SQ SEQUENCE 937 AA; 106275 MW; 6A775209C305DFE5 CRC64;

Query Match 10.9%; Score 111; DB 1; Length 937;
Best Local Similarity 21.0%; Pred. No. 0.051;
Matches 43; Conservative 45; Mismatches 73; Indels 44; Gaps 9;

QY 1 IAKEVSTERTYKLDLEVTISWFQSTVSKEDAMPEAL-----KSLIFPNFELHKEHT 53
Db 511 VMNELDTERAYVEELLCVLEGYAAEMD-NPLMAHLISTGLQNKKNILFGNMEEIYHFHN 569
QY 54 N-FLKEIEQRLALWE--GRSNAQIRDYQRIQDVMLKNIQGMKHLAAHLWKHSEALEALEN 110
Db 570 RIFLRELESCIDCPELYGRCFELERMEEFOIYEKYCN-----KPRSESLNRQ----- 616
QY 111 GIKSSRLNFCRDFELQKVY-----LPLNTFLRLPLHMLHYKOVLERLCKKHPPSH 164
Db 617 -----CSDCPFFQECQKKLDHKLSDSYLLKPVQIRITKYQLLLKMLKY--SKH 663
QY 165 AD-FRDCRAALAEITENVAOLHGTM 188
Db 664 CEGAEDLQEALESILGILKAVNDMSM 688

```

Search completed: December 6, 2001, 08:53:10
Job time: 382 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:54:32 ; Search time 77.11 Seconds
 (without alignments)
 368.005 Million cell updates/sec

Title: US-09-555-342a-2_COPY_544_737
 Perfect score: 1023
 Sequence: 1 IAKEVSTERYLKDELVIT.....AEITEMVAQLHGTMIKMF 194

Scoring table:
 BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL17.*
 1: sp-archaea.*
 2: sp-bacteria.*
 3: sp-fungi.*
 4: sp-human.*
 5: sp-invertebrate.*
 6: sp-mammal.*
 7: sp-mhc.*
 8: sp-organelle.*
 9: sp-phage.*
 10: sp-plant.*
 11: sp-rodent.*
 12: sp-virus.*
 13: sp-vertebrate.*
 14: sp-unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1023	100.0	1045	4 Q9Y4F1	Q9Y4F1 homo sapien
2	570	55.7	420	11 Q99KU3	Q99KU3 mus musculus
3	566.5	55.4	1054	4 Q94887	Q94887 homo sapien
4	320	31.3	562	5 Q9VN69	Q9VN69 drosophila
5	195.5	19.1	766	11 Q88387	Q88387 rattus norv
6	190.5	18.6	875	5 Q18372	Q18372 caenorhabdi
7	184.5	18.0	699	4 Q9P215	Q9P215 homo sapien
8	179	17.5	1676	4 Q9ULG4	Q9ULG4 homo sapien
9	179	17.5	1681	4 Q9NYG0	Q9NYG0 homo sapien
10	179	17.5	1696	4 Q9NZM3	Q9NZM3 homo sapien
11	176	17.2	411	11 Q9ER22	Q9ER22 rattus norv
12	176	17.2	493	11 Q9X73	Q9X73 rattus norv
13	176	17.2	516	4 Q43307	Q43307 homo sapien
14	172	16.8	1721	4 Q9S216	Q9S216 homo sapien
15	171	16.7	496	4 Q9NTG0	Q9NTG0 homo sapien
16	171	16.7	619	4 Q9HDC6	Q9HDC6 homo sapien
17	171	16.7	621	13 Q93504	Q93504 brachydanio
18	171	16.7	694	4 Q9UPP0	Q9UPP0 homo sapien
19	171	16.7	720	4 Q9NR80	Q9NR80 homo sapien

20	168.5	16.5	1658	11 Q9Z0R6	Q9Z0R6 mus musculus
21	166.5	16.3	1335	5 Q00886	Q00886 dictyosteli
22	166	16.2	820	5 Q96035	Q96035 ciona savig
23	166	16.2	1721	4 Q9UNK2	Q9UNK2 homo sapien
24	163	15.9	1714	11 Q9Z0R4	Q9Z0R4 mus musculus
25	157.5	15.4	1737	5 Q9TW28	Q9TW28 dictyosteli
26	157	15.3	698	4 Q9NXY1	Q9NXY1 homo sapien
27	155	15.2	727	11 Q88841	Q88841 mus musculus
28	154	15.1	647	5 Q9VS45	Q9VS45 drosophila
29	152	14.9	733	11 Q88842	Q88842 mus musculus
30	149	14.6	1522	4 Q15085	Q15085 homo sapien
31	147	14.4	1527	11 Q9ES67	Q9ES67 rattus norv
32	138	13.5	761	3 Q9HF59	Q9HF59 ashbaya goss
33	131	12.8	1544	4 Q9NZN5	Q9NZN5 homo sapien
34	129.5	12.7	985	11 Q9ESG7	Q9ESG7 mus musculus
35	129	12.6	618	11 Q9CWR0	Q9CWR0 mus musculus
36	128.5	12.6	1028	4 Q9UGQ4	Q9UGQ4 homo sapien
37	128	12.5	1101	5 Q9VEF9	Q9VEF9 drosophila
38	127.5	12.5	1011	5 Q9W0R9	Q9W0R9 drosophila
39	127.5	12.5	1011	5 Q9GTU9	Q9GTU9 drosophila
40	126.5	12.4	750	4 Q15086	Q15086 homo sapien
41	125.5	12.3	958	4 Q9H023	Q9H023 homo sapien
42	125	12.2	594	5 Q9VS95	Q9VS95 drosophila
43	124.5	12.2	793	5 Q9NHV9	Q9NHV9 drosophila
44	123.5	12.1	852	5 Q9VWJ5	Q9VWJ5 drosophila
45	123	12.0	630	11 Q9ES27	Q9ES27 mus musculus

ALIGNMENTS

RESULT 1
 Q9Y4F1 ID Q9Y4F1 PRELIMINARY; PRT; 1045 AA.
 AC Q9Y4F1;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CDEP.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CARTILAGE;
 RX MEDLINE=98086358; PubMed=9425278;
 RA Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,
 RA Kato Y.;
 RT "Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors.";
 RL Biochem. Biophys. Res. Commun. 241:369-375(1997).
 DR EMBL; AB008430; BAA24267.1; -
 DR HSSP; P08567; iPLS.
 DR InterPro; IPR000299; Band_4.1.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; RhoGEF.
 DR Pfam; PF00373; Band_41; 1.
 DR Pfam; PF00169; PH; 2.
 DR Pfam; PF00621; RhoGEF; 1.
 DR PRINTS; PR00935; BAND41.
 DR SMART; SM00295; B41; 1.
 DR SMART; SM00233; PH; 2.
 DR SMART; SM00325; RhoGEF; 1.
 DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
 DR PROSITE; PS50037; BAND_41_3; 1.
 DR PROSITE; PS50003; PH_DOMAIN; 2.
 SQ SEQUENCE 1045 AA; 118632 MW; 0E8B2D61C0F58417 CRC64;

Query Match 100.0%; Score 1023; DB 4; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 1.4e-81;

```
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IAEVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 60
DB 544 IAEVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 603
QY 61 QRLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLEN 120
DB 604 QRLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLEN 663
QY 121 FCRDFELQKVCYLPNTFLRLPHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITEM 180
DB 664 FCRDFELQKVCYLPNTFLRLPHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITEM 723
QY 181 VAQLHGTMIKMF 194
DB 724 VAQLHGTMIKMF 737

RESULT 2
Q99KU3 ID Q99KU3 PRELIMINARY; PRT; 420 AA.
AC Q99KU3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE UNKNOWN (PROTEIN FOR IMAGE:3493093) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004009; AAH04009.1; -.
FT NON_TER 1
SQ SEQUENCE 420 AA; 48387 MW; 804411A6E6613F96 CRC64;

Query Match 55.7%; Score 570; DB 11; Length 420;
Best Local Similarity 94.8%; Pred. No. 3e-42;
Matches 106; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 83 VMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCRDFELQKVCYLPNTFLRLP 142
DB 1 VMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCRDFELQKVCYLPNTFLRLP 60
QY 143 LHRMHYKQVLERLCKHPPSHADFRDCRAALAEITEMVAQLHGTMIKMF 194
DB 61 LHRMHYKQVLERLCKHPPSHADFRDCRAALAEITEMVAQLHGTMIKMF 112

RESULT 3
Q94887 ID Q94887 PRELIMINARY; PRT; 1054 AA.
AC Q94887;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE KIAA0793 PROTEIN.
GN KIAA0793.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
```

```
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro. ";
RL DNA Res. 5:277-286(1998).
RN [2]
RP SEQUENCE OF 1-754 FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence. ";
RL Genome Res. 8:1097-1108(1998).
RN [3]
RP SEQUENCE OF 1-754 FROM N.A.
RA Sun H., Stoneking T., Langston Y., Laplant Y.;
RT "The sequence of Homo sapiens BAC clone RG442F18. ";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-754 FROM N.A.
RA Waterston R.H.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-754 FROM N.A.
RA Waterston R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 1-754 FROM N.A.
RA Waterston R.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB018336; BAA34513.1; -.
DR EMBL; AC005104; AAD12224.1; -.
DR HSSP; P08567; IPLS.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001849; PH.
DR InterPro; IPR00219; RhogEF.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00169; PH; 2.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhogEF; 1.
DR PROSITE; PS00660; BAND_41.1; UNKNOWN_1.
DR PROSITE; PS50057; BAND_41_3; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
SQ SEQUENCE 1054 AA; 119887 MW; 7DC279F69A307E5A CRC64;

Query Match 55.4%; Score 566.5; DB 4; Length 1054;
Best Local Similarity 53.1%; Pred. No. 1.7e-41;
Matches 103; Conservative 36; Mismatches 54; Indels 1; Gaps 1;
QY 1 IAEVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 60
DB 539 IAEVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 598
QY 61 QRLALWEGRSNAQIR-DYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 119
DB 599 QRLALWEGRSNAQIR-DYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 658
QY 120 NFCRDFELQKVCYLPNTFLRLPHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITE 179
DB 659 AVYKFEFELQKVCYLPNTFLRLPHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITE 718
QY 180 MVAQLHGTMIKMF 193
DB 719 VTTTLQHILIRLEN 732

RESULT 4
Q9VN69 ID Q9VN69 PRELIMINARY; PRT; 562 AA.
AC Q9VN69;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
```

CG2008 PROTEIN.

CG2008.

GN Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OC NCBI_TAXID=7227;

LN [1]

RN SEQUENCE FROM N.A.

RP STRAIN=BERKELEY;

RC MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananatzides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Aqbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Boudreau N., Brothier P.,

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., D.L.,

RA Palazzolo M., Pittman G.S., Pan S., Saunders R.D., Reese M.G.,

RA Reinert K., Remington K., Sanders P.O., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.-S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RT Science 287:2185-2195(2000).

RL EMBL: AE003604; RAF52080.1; -.

DR HSSP: P08567; IPLS.

DR Flybase; FBgn0037287; CG2008.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000219; RhoGEF.

DR Pfam; PF00169; PH; 2.

DR Pfam; PF00621; RhoGEF; 1.

DR SMART; SM00233; PH; 1.

DR SMART; SM00325; RhoGEF; 1.

DR PROSITE; PSS0003; PH_DOMAIN; 1.

SQ SEQUENCE 562 AA, 64296 MW; D5470EB20E3C91B9 CRC64;

```
Query Match      31.3%; Score 320; DB 5; Length 562;
Best Local Similarity 38.4%; Pred. No. 3.6e-20;
Matches         66; Conservative 35; Mismatches 65; Indels 6; Gaps 3;
```

```

Qy 121 FCRDFELQKVCYLPNTFLRLPHRLMHYKQVLERLCKHHPPSHADFRDRA 172
      ::|| ||||||| :||:||||: :||||: | :| :| :|
Db 209 VYKEFQQKCVCYLPIGELLGLPLNRLHYQLILERCLDYGEEDHIDYADAMA 260

RESULT 5
O88387 PRELIMINARY; PRT; 766 AA.
AC O88387;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ACTIN-FILAMENT BINDING PROTEIN FRABIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98334590; PubMed=9668039;
RA Obaishi H., Nakanishi H., Mandai K., Satoh K., Satoh A., Takahashi K.,
RA Miyahara M., Nishiooka H., Takaishi K., Takai Y.;
RT "Frabin, a novel FGD1-related actin filament-binding protein capable
RT of changing cell shape and activating c-Jun N-terminal kinase.";
RL J. Biol. Chem. 273:18697-18700(1998).
DR EMBL; AF038388; AAC27698.1; -.
DR InterPro; IPR000219; RhogEF.
DR InterPro; IPR000306; Znf_FYVE.
DR InterPro; IPR000822; Znf-C2H2.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00621; RhogEF; 1.
DR Pfam; PF01363; FYVE; 1.
DR PROSITE; PS00003; PH_DOMAIN; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhogEF; 1.
SQ SEQUENCE 766 AA; 86453 MW; E3483DC0B0B334C CRC64;

Query Match 19.1%; Score 195.5; DB 11; Length 766;
Best Local Similarity - 26.3%; Pred. No 4.3e-09;
Matches 5; Conservative 43; Mismatches 90; Indels 13; Gaps

Qy 1 TAKEVSTERTYTKDLGVITSWFSQTSYKE---DAMPEALKSLIFPNFELPKHFHTNF-L 56
      || : ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 210 IATELLTERAYVSRLNDQVFYCKLEEANRGSPFAEMVNKIFSNISSINAFHSKFL 269

Qy 57 KEIQRIALWEGRNAQIRDYQRIGDVMLKNIOGMKHLAHLMKHSEALENGIKSSR 116
      |:|:|: || |||||:| :| :| :| :| :| :| :| :| :| :| :| :|
Db 270 PELSKRWQENE-----TPPRIGDIQLKLAPFLKMVGYYVKGFDNAVELVKNMTERVP 321

Qy 117 RLENFCRDFELQKVC-YLPLNTFLRLPHRLMHYKQVLERLCKHHPPSHADFRDRAALA 175
      : : : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 322 QFKSVTEIQKQICGSLTQHMHLEFIQRIPTYEMLLKDKYLKKLSPADPDWDNAKSLE 381

Qy 176 EITEWVAQLHGTMTKMEN 193
      | : : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 382 IIISTAASHNSAIRKMEN 399

RESULT 6
I18372 PRELIMINARY; PRT; 875 AA.
AC Q18372;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE C33D9.1 PROTEIN.
OS Caenorhabditis elegans.

```

```
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd C.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smauld N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RL EMBL; Z68159; CAA92282.1; -.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000306; Znf_FYVE.
DR InterPro: IPR001849; PH.
DR Pfam; PF00169; PH; 2.
DR SMART; SM00064; FYVE; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
DR Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RL EMBL; Z68159; CAA92282.1; -.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000306; Znf_FYVE.
DR InterPro: IPR001849; PH.
DR Pfam; PF00169; PH; 2.
DR SMART; SM00064; FYVE; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhoGEF; 1.
SQ SEQUENCE 875 AA; 100290 MW; 2C7DEA1EE2C25852 CRC64;

Query Match 18.6%; Score 190.5; DB 5; Length 875;
Best Local Similarity 25.6%; Pred. No. 1.4e-08;
Matches 50; Conservative 46; Mismatches 88; Indels 11; Gaps 4;

QY 2 AKEVSTTERTYRKDLVITSWFQSTVSKEDAMPEALKSLFFPEPLKHH-TNFKLEIE 60
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 AKELVDSQRYVDKLLGDTFRNLRIKEELITNDKTRLLANVSSLYQFNTHFLPQL 392
QY 61 QRLALWEGRSNAQIRDYQIRGDMVKMLKLAHLWKHSEALENGIKSSRLN 120
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 393 ESIRDW-----HTTKRIANVVRKQAPFLKMYSETNNYDRATKLFEE-LKKKKFAD 443
QY 121 FCRDFELQKVCY-LPLNTFLRLPLHRLMHYKQVLERCKHPPSHADFRDCRAALAEITE 179
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 444 VKKEIKQAECEGLPLGHLLICPVQVMRYQLLLQEVKKHLQPSDVFDDTTVALELVQ 503
QY 180 MVAQLHGTMIKMFN 194
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 504 AAHANEMMKLDRF 518

RESULT 7
Q9P215 PRELIMINARY; PRT; 699 AA.
AC Q9P215
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA1362 PROTEIN (FRAGMENT).
GN KIAA1362.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
```

```
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037783; BAA92600.1; -.
DR InterPro: IPR001304; Lectin_c.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PSS00615; C-TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE; PSS0003; PH_DOMAIN; 2.
FT NON_TER 1
SQ SEQUENCE 699 AA; 79952 MW; 3EFBA42718F07C80 CRC64;

Query Match 18.0%; Score 184.5; DB 4; Length 699;
Best Local Similarity 25.1%; Pred. No. 3.6e-08;
Matches 52; Conservative 47; Mismatches 81; Indels 27; Gaps 4;

QY 1 IAKEVSTTERTYRKDLVITSWFQSTVSKEDAMPEALKSLFFNFE-----PL 48
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 IAKEIMSSSEKVFVDVLKLLHIDF-----DAVAHASRQLGKGVIEDRIILYLPQL 260
QY 49 HKFHTNFKLEIEQRLALWEGRSNAQIRDYQIRGDMVKMLKLAHLWKHSEALEAL 108
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 YELNRDLKLEERMLHW-----TEQRIADIFVKKGPKLKMYSYIKFEDKNIAL 312
QY 109 ENGIKSSRLNFCRDFELQKVC-YLPLNTFLRLPLHRLMHYKQVLERCKHPPSHADF 167
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 DEQCKNPGFAAVVREFEMSPCCANLAKHLLKPVQIRPQYRLLLTDLKLNIEDAGDY 372
QY 168 RDCRAALAEITEVMAQLHGTMIKMFN 194
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 RDTQDALAVVIEVANHANDTMKQDNF 399

RESULT 8
Q9ULG4 PRELIMINARY; PRT; 1676 AA.
AC Q9ULG4
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA1256 PROTEIN (FRAGMENT).
GN KIAA1256.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
CC -I- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL; AB033082; BAA86570.1; -.
DR HSP; P06241; 1SHF.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR000108; Neu_cyt_fact_2.
```

```
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR Pfam: PF00018; SH3; 5.
DR PRINTS: PR00499; P67PHOX.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM00054; EFH; 2.
DR SMART: SM00027; EH; 2.
DR SMART: SM00325; RhoGEF; 1.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS00004; C2_DOMAIN_2; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50002; SH3; 5.
KW Calcium-binding.
FT NON_TER 1
SQ SEQUENCE 1676 AA; 191178 MW; B877291582FB7F99 CRC64;

Query Match 17.5%; Score 179; DB 4; Length 1676;
Best Local Similarity 25.8%; Pred. No. 2.9e-07;
Matches 51; Conservative 47; Mismatches 80; Indels 20; Gaps 5;

QY 4 EVSTTERTYKDLVITSFQSTVSKEDAMPEALKSLIFPNFELPKHFTNFKLEIQRL 63
DB 1195 ELIQTEERYMADQLVVEVFKRMAESGFLTEGEMALIFVNMKELIMSNKLLKALVR- 1253

QY 64 ALWEGSRNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCR 123
DB 1254 ----KTGGEKMPVQMIGDILAAE---LSHMQAYIRFCSCQL----NGAALQOKTDEDT 1302

QY 124 DFE--LQKVCY-----LPLNTFLLRPLRLMHYKOVLERLCKHPPSHADFRCDRAALA 175
DB 1303 DKFEFLKLKASDPRCKGMPLSSFLPKPMQIRITRYPLLRISILENTPESHADHSSKLAL 1362

QY 176 EITEMVAQLHGTMKMN 193
DB 1363 RAEELCSQVNEGVRKEN 1380

RESULT 9
Q9NYG0
ID Q9NYG0 PRELIMINARY; PRT; 1681 AA.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INTERSECTIN 2.
GN SH3DLB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Seifert M., Engel M., Welter C.;
RT "Intersectin 2 (SH3DLB), human homolog of mouse Ecs2 protein.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AF248540; AAF63600.1; -.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR000108; Neu_cyt_fact_2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.

DR InterPro: IPR00168; C2; 1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR PRINTS: PR00499; P67PHOX.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM00054; EFH; 2.
DR SMART: SM00027; EH; 2.
DR SMART: SM00325; RhoGEF; 1.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS00004; C2_DOMAIN_2; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50002; SH3; 5.
KW Calcium-binding.
FT NON_TER 1
SQ SEQUENCE 1676 AA; 191178 MW; B877291582FB7F99 CRC64;

Query Match 17.5%; Score 179; DB 4; Length 1681;
Best Local Similarity 25.8%; Pred. No. 2.9e-07;
Matches 51; Conservative 47; Mismatches 80; Indels 20; Gaps 5;

QY 4 EVSTTERTYKDLVITSFQSTVSKEDAMPEALKSLIFPNFELPKHFTNFKLEIQRL 63
DB 1200 ELIQTEERYMADQLVVEVFKRMAESGFLTEGEMALIFVNMKELIMSNKLLKALVR- 1258

QY 64 ALWEGSRNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCR 123
DB 1259 ----KTGGEKMPVQMIGDILAAE---LSHMQAYIRFCSCQL----NGAALQOKTDEDT 1307

QY 124 DFE--LQKVCY-----LPLNTFLLRPLRLMHYKOVLERLCKHPPSHADFRCDRAALA 175
DB 1308 DKFEFLKLKASDPRCKGMPLSSFLPKPMQIRITRYPLLRISILENTPESHADHSSKLAL 1367

QY 176 EITEMVAQLHGTMKMN 193
DB 1368 RAEELCSQVNEGVRKEN 1385

RESULT 10
Q9NZM3
ID Q9NZM3 PRELIMINARY; PRT; 1696 AA.
AC Q9NZM3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INTERSECTIN 2 LONG ISOFORM.
GN ITS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pucharcos C., Estivill E., de la Luna S.;
RT "Cloning and characterization of human intersectin 2: a protein
involved in endocytosis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AF182198; AAF59903.1; -.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR000108; Neu_cyt_fact_2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00169; PH; 1.
```

```

DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 5.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00239; C2; 1.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00002; SH3; 5.
DR Calcium-binding.
KW SEQUENCE 1696 AA; 193325 MW; 17B5CDAC9BBCF73B CRC64;

Query Match 17.5%; Score 179; DB 4; Length 1696;
Best Local Similarity 25.8%; Pred. No. 2.9e-07;
Matches 51; Conservative 47; Mismatches 80; Indels 20; Gaps 5;

QY 4 EVSTERTYLYKDL-EVITSWFQSTVSKEDAMPEALKSLIFNFPFLKHFHTNFKLKEI 63
Db 1215 ELIQTEERYMADLQLVVEVFQKRMAESGFTLEGEMALIFVNWKELIMSNTKLLKALRVR- 1273
QY 64 ALWEGRSNAQIRDYQRIQIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCR 123
Db 1274 ----KKTGGEKMPVQMGIDILAE---LSHQAYIRCSQQL-----NGAALLQOKTDDET 1322
QY 124 DFE--LQKVCY-----LPLNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRDCRAALA 175
Db 1323 DFEFLKKLASDPKCGMPLSSFLKPKMQIRTRYPLLRISILENTPESHADHSLKLALE 1382
QY 176 EITMVAQLHGTMIKMN 193
Db 1383 RAEELCSQVNEGVRKEN 1400

RESULT 11
Q9ER22 PRELIMINARY; PRT; 411 AA.
AC Q9ER22;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE COLLYBISTIN II.
GN COLLYBISTIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20075283; PubMed=10607391;
RA Kins S., Betz H., Kirsch J.;
RT "Collybistin, a newly identified brain-specific GEF, induces
RT submembrane clustering of gephyrin.";
RL Nat. Neurosci. 3:22-29(2000).
DR EMBL; AJ250425; CAB65966.1; -.
DR HSSP; Q07889; IAWF.
DR InterPro; IPR001849; PH.
DR InterPro; IPR00219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00002; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
RP SEQUENCE 493 AA; 58157 MW; 41040671B9398BFA CRC64;

Query Match 17.2%; Score 176; DB 11; Length 493;
Best Local Similarity 25.0%; Pred. No. 1.4e-07;
Matches 49; Conservative 43; Mismatches 92; Indels 12; Gaps 4;

QY 1 IAEVSTERTYLYKDL-EVITSWFQSTVSKEDAMPEALKSLIFNFPFLKHFHTNFKLKEI 59
Db 114 VINEIMSTERHYIKHLKDICEGYLKQCRKRDMSFSDQLKVIQGNIEDIYRFQMGFYRDL 173
QY 60 EORLALWEGRSNAQIRDYQRIQIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 119
Db 174 EKQY-----NNDPDLHSLSEIGCFLEHQGFWIYSEYCNHLDACMELSKLMKDSRYQH 226
QY 120 NF--CRDFELQKVCYLPNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRDCRAALAEI 177
Db 227 FFEACR--LLQQMDIDTIDGFLTPVQKICKYPLQLAELLKYTAQDHSYRYVAALAVM 284
QY 178 TEMVAQLHGTMIKMN 193
Db 178 TEMVAQLHGTMIKMN 193

Query Match 17.2%; Score 176; DB 11; Length 411;
Best Local Similarity 25.0%; Pred. No. 1.1e-07;
Matches 49; Conservative 43; Mismatches 92; Indels 12; Gaps 4;

```

```

QY 1 IAEVSTERTYLYKDL-EVITSWFQSTVSKEDAMPEALKSLIFNFPFLKHFHTNFKLKEI 59
Db 54 VINEIMSTERHYIKHLKDICEGYLKQCRKRDMSFSDQLKVIQGNIEDIYRFQMGFYRDL 113
QY 60 EORLALWEGRSNAQIRDYQRIQIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 119
Db 114 EKQY-----NNDPDLHSLSEIGCFLEHQGFWIYSEYCNHLDACMELSKLMKDSRYQH 166
QY 120 NF--CRDFELQKVCYLPNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRDCRAALAEI 177
Db 167 FFEACR--LLQQMDIDTIDGFLTPVQKICKYPLQLAELLKYTAQDHSYRYVAALAVM 224
QY 178 TEMVAQLHGTMIKMN 193
Db 225 RNVTQQLNKRRLN 240

RESULT 12
Q9QX73 PRELIMINARY; PRT; 493 AA.
AC Q9QX73;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE COLLYBISTIN I.
GN COLLYBISTIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20075283; PubMed=10607391;
RA Kins S., Betz H., Kirsch J.;
RT "Collybistin, a newly identified brain-specific GEF, induces
RT submembrane clustering of gephyrin.";
RL Nat. Neurosci. 3:22-29(2000).
DR EMBL; AJ250425; CAB65966.1; -.
DR HSSP; Q07889; IAWF.
DR InterPro; IPR001849; PH.
DR InterPro; IPR00219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00002; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
RP SEQUENCE 493 AA; 58157 MW; 41040671B9398BFA CRC64;

```


RC	TISSUE=BRAIN;									
RX	MEDLINE=99017974; PubMed=9799604;									
RA	Giulipponi M., Scott H.S., Chen H.A., Rossier C.,									
RA	Antonarakis S.E.;									
RT	Two isoforms of a human intersectin (ITSN) protein are produced by									
RT	brain-specific alternative splicing in a stop codon.";									
CC	Genomics 53:369-376(1998).									
CC	-1- SIMILARITY: TO EF-HAND FAMILY.									
DR	EMBL: AF064244; AAC78611.1; -.									
DR	HSSP: P29355; 1SEM.									
DR	InterPro: IPR000008; C2.									
DR	InterPro: IPR002048; EF-hand.									
DR	InterPro: IPR000261; EPS15_repeat.									
DR	InterPro: IPR000108; Neu_cyt_fact_2.									
DR	InterPro: IPR001849; PH.									
DR	InterPro: IPR000219; RhoGEF.									
DR	InterPro: IPR001452; SH3.									
DR	Pfam: PF00168; C2; 1.									
DR	Pfam: PF00036; ehand; 3.									
DR	Pfam: PF00169; PH; 1.									
DR	Pfam: PF00621; RhoGEF; 1.									
DR	Pfam: PF00018; SH3; 5.									
DR	PRINTS: PR00499; P67PHOX.									
DR	PRINTS: PR00452; SH3DOMAIN.									
DR	SMART: SM00239; C2; 1.									
DR	SMART: SM00054; EFh; 2.									
DR	SMART: SM00027; EH; 2.									
DR	SMART: SM00233; PH; 1.									
DR	SMART: SM00325; RhoGEF; 1.									
DR	SMART: SM00326; SH3; 5.									
DR	PROSITE: PS50004; C2_DOMAIN_2; 1.									
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_2.									
DR	PROSITE: PS50003; PH_DOMAIN; 1.									
DR	PROSITE: PS50002; SH3; 5.									
KW	Calcium-binding.									
SQ	SEQUENCE 1721 AA; 195559 MW; 851A5CFB2BC4EBFC CRC64;									
Query Match 16.8%; Score 172; DB 4; Length 1721;										
Best Local Similarity 20.9%; Pred. No. 1.2e-06;										
Matches 40; Conservative 52; Mismatches 93; Indels 6; Gaps										
Qy	4	EVSTTERTYIKLEVITVSFQTSVKEDAMPALKSLIPNFPDLPHKHFNFILKETQRL	63							
Db	1244	ELIVTEENVNDLQLVTEIFQKPLMESELLTEKEVAMIFVNWKELINCNILKALKRVR-	1302							
Qy	64	ALWGRSNAQIRDYQIRIGDVMKLNQGMKHLAAHLWKHSEALELENGIKSSRRLENFCR	123							
Db	1303	-----KMSGKMKVNMIGDILSAQLPHQPIRFGCSRLNGAALIQKTEADPDFEVK	1358							
Qy	124	DFELQKVCY-LPLNTFLRLPHRLMHYQVLERLCKRHPPSHADFCRAALAEITEMVA	182							
Db	1359	RLEMDPRCKMPLSSFLKPMQVRVTRYPLIINKNILENTPEHDPHSHLKHALEKAELCS	1418							
Qy	183	QLHGTMIKMN	193							
Db	1419	QVNEGVRKEN	1429							
RESULT 15										
Q9NTG0										
ID	Q9NTG0	PRELIMINARY;	PRT;	496	AA.					
AC	Q9NTG0;									
DT	01-OCT-2000	(TrEMBLrel. 15, Created)								
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)								
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)								
DE	HYPOTHEtical 57.7 KDA PROTEIN (FRAGMENT).									
GN	DKF2P434G2016.									
OS	Homo sapiens (Human).									
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;									
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.									
OX	NCBI_Taxid=9606;									
RN	[1]									

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:51:40 ; Search time 72.4 Seconds
(without alignments)
382.644 Million cell updates/sec

Title: US-09-555-342A-2_COPY_1_374

Perfect score: 1956

Sequence: 1 MGEIERPTGSLGAPENS.....GHKKVQFERKSHIRSL 374

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseqp/AA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1956	100.0	1045	20 AAY07482	Human chondrocyte-
2	1952	99.8	1045	21 AAY91947	Human cytoskeleton
3	689.5	35.3	1005	21 AAY91946	Human cytoskeleton
4	587.5	30.0	687	22 AAB95258	Human protein sequ
5	506.5	25.9	913	18 AAW12522	Protein tyrosine p
6	506.5	25.9	913	20 AAY25156	Human TPPI1 protei
7	399.5	20.4	369	22 AAU04486	Human protein tyro
8	359	18.4	635	21 AAB53356	Human colon cancer
9	356	18.2	1174	19 AAW67438	Human protein tyro
10	356	18.2	1174	20 AAY34158	Human protein tyro
11	348	17.8	586	20 AAY27443	Amino acid sequenc

12	328	16.8	590	20 AAW94458	Human neurofibroma
13	328	16.8	591	18 AAW09648	Human merlin prote
14	328	16.8	595	15 AAR60398	Merlin protein enc
15	328	16.8	596	20 AAW94459	Human neurofibroma
16	327	16.7	584	20 AAW94457	Mouse neurofibroma
17	327	16.7	591	20 AAW94456	Mouse neurofibroma
18	327	16.7	596	20 AAW94455	Mouse neurofibroma
19	326.5	16.7	305	22 AAM25892	Human protein sequ
20	325	16.6	584	18 AAW09647	Mouse merlin prote
21	323	16.5	591	18 AAW09646	Mouse merlin prote
22	323	16.5	596	18 AAW09645	Mouse merlin prote
23	320.5	16.4	495	22 AAM41770	Human polypeptide
24	312	16.0	445	22 AAM33984	Human polypeptide
25	286.5	14.6	279	22 AAM25733	Human protein sequ
26	266	13.6	450	22 AAB92607	Human protein sequ
27	265	13.5	2466	16 AAR71498	Human protein tyro
28	265	13.5	2466	19 AAW75999	Intracellular prot
29	265	13.5	2466	21 AAY90272	Human TPPI1 phosph
30	265	13.5	2485	21 AAB19343	Amino acid sequenc
31	245	12.5	142	22 AAG74414	Human colon cancer
32	223.5	11.4	1105	22 AAM25567	Human protein sequ
33	166	8.5	1007	22 AAB50660	C. elegans UNC-5 p
34	164	8.4	184	22 AAM25248	Human protein sequ
35	157	8.0	436	22 AAG73954	Human colon cancer
36	141	7.2	91	22 AAM13919	Peptide #353 encod
37	141	7.2	91	22 AAM26325	Peptide #362 encod
38	141	7.2	91	22 AAM01662	Peptide #344 encod
39	128	6.5	92	21 AAG01312	Human secreted pro
40	124	6.3	94	21 AAG01151	Human secreted pro
41	123.5	6.3	348	21 AAB42166	Human ORFX ORF1930
42	121.5	6.2	91	21 AAG00826	Human secreted pro
43	119	6.1	97	21 AAB56466	Human prostate can
44	115	5.9	2541	21 AAB41087	Human ORFX ORF851
45	112	5.7	69	22 AAM36561	Peptide #10598 enc

ALIGNMENTS

RESULT 1
AAY07482
ID AAY07482 standard; Protein; 1045 AA.

XX AC AAY07482;

XX DT 17-AUG-1999 (first entry)

XX DE Human chondrocyte-derived protein CDEP.

XX KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
XX KW Db1 homology domain; plectstrin homology domain; rheumatoid 'arthritis;
XX KW drug.

XX OS Homo sapiens.

XX PN WO9928458-A1.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-JP05348.

XX PR 27-NOV-1997; 97JP-0342060.

XX PA (CHUS) CHUGAI SEIYAKU KK.

XX PI Kato Y, Kawamoto T, Koyano Y;

XX DR WPI; 1999-371117/31.

XX DR N-PSDB; AAX79183.

XX PT Protein CDEP expressed in differentiated chondrocytes, and gene

XX encoding it

XX

Modified-site	878	/note=	"potential phosphorylation site"
Modified-site	883	/note=	"potential phosphorylation site"
Modified-site	889	/note=	"potential phosphorylation site"
Modified-site	898	/note=	"potential phosphorylation site"
Modified-site	919	/note=	"potential phosphorylation site"
Modified-site	923	/note=	"potential N-glycosylation site"
Modified-site	966	/note=	"potential phosphorylation site"
Modified-site		/note=	"potential phosphorylation site"
WO200017355-A2.			
30-MAR-2000.			
17-SEP-1999;	99WO-US21565.		
18-SEP-1998;	98US-0172226.		
27-APR-1999;	99US-0131321.		
(INCY-) INCYTE PHARM INC.			
Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;			
Guegler KJ, Patterson C, Azimzai Y, Baughn MR;			
WPI; 2000-283582/24.			
N-PSDB; AAA08582.			
Human cytoskeleton associated proteins, used to treat cell			
proliferative, autoimmune/inflammatory, vesicle trafficking,			
neurological, cell motility, reproductive and muscle disorders			
Claim 1; Page 82-84; 113pp; English.			
AAV91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat and			
diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle			
trafficking, neurological, cardiovascular, cell motility, reproductive			
and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to			
CYSKP-16 can be used to treat or prevent disorders associated with			
decreased expression or activity of CYSKP (claimed), for example,			
atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,			
cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,			
asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,			
diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma			
and trauma. CYSKP antagonists can be used to treat or prevent a			
disorder associated with increased expression or activity of CYSKP			
(claimed).			
Sequence 1045 AA;			
Query Match 99.8%; Score 1952; DB 21; Length 1045;			
Best Local Similarity 99.7%; Pred. No. 2.7e-190;			
Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0			
Qy 1 MGEIEQPTPGSRIGAPENSIGISTLERGQKPPPTSGKLIVSIKIQMLDDBTQFAFVPPQRA 60			
1 mgeleqrtpgsgirgapensgistlergqkppptsgklivsiqimldtqeaefvqra 60			
Qy 61 PGKVLLDAVNCNHLNIVEGDYFGLFPPDHKKITVWLDLLKPIVKQIRRPKHVVVKFVKKFF 120			
Db 61 pgkvlldavcnhnlivegdylfepdhkkitvwdllkplvkqirrpkhvvkvfkvff 120			
Qy 121 PPDHQLQELTRYLFALOVKQDLAQLTQNDTSAALLSHIVQSEITGCDPEALDRHL 180			
Db 121 ppdhtqlqeitrylfaqlvqkqlagrltqndtsaallshivqseigtdpealdrhl 180			
Qy 181 AKNKYIPQDALEDKIVFEHNNHIGQTPAESDFQLEIARRLWEMTGIRLHPAKDREGTKI 240			

FT Modified-site 449 /note= "potential phosphorylation site"

FT Modified-site 518 /note= "potential phosphorylation site"

FT Modified-site 530 /note= "potential phosphorylation site"

FT Modified-site 543 /note= "potential phosphorylation site"

FT Modified-site 544 /note= "potential phosphorylation site"

FT Modified-site 548 /note= "potential phosphorylation site"

FT Modified-site 614 /note= "potential phosphorylation site"

FT Modified-site 623 /note= "potential phosphorylation site"

FT Modified-site 647 /note= "potential phosphorylation site"

FT Modified-site 658 /note= "potential phosphorylation site"

FT Modified-site 673 /note= "potential phosphorylation site"

FT Modified-site 682 /note= "potential phosphorylation site"

FT Domain 718..721 /note= "signature sequence"

FT Modified-site 730 /note= "potential phosphorylation site"

FT Modified-site 744 /note= "potential phosphorylation site"

FT Modified-site 746 /note= "potential phosphorylation site"

FT Modified-site 748 /note= "potential phosphorylation site"

FT Modified-site 766 /note= "potential phosphorylation site"

FT Modified-site 828 /note= "potential phosphorylation site"

FT Modified-site 854 /note= "potential phosphorylation site"

FT Modified-site 879 /note= "potential phosphorylation site"

FT Modified-site 884 /note= "potential phosphorylation site"

FT Modified-site 890 /note= "potential phosphorylation site"

FT Modified-site 944 /note= "potential phosphorylation site"

FT Modified-site 952 /note= "potential phosphorylation site"

FT Modified-site /note= "potential phosphorylation site"

WO200017355-A2.

PD 30-MAR-2000.

XX 17-SEP-1999; 99WO-0521565.

XX 18-SEP-1998; 98US-0172226.

PR 27-APR-1999; 99US-0131321.

XX (INCY-) INCYTE PHARM INC.

XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;

PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;

XX WPI; 2000-283582/24.

DR N-PSDB; AAA08581.

XX Human cytoskeleton associated proteins, used to treat cell

PT proliferative, autoimmune/inflammatory, vesicle trafficking,

PT neurological, cell motility, reproductive and muscle disorders

XX

PS Disclosure; Page 79-82; 113pp; English.

XX AAY91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1

CC to CYSKP-16) respectively. The sequences can be used to treat and

CC diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle

CC trafficking, neurological, cardiovascular, cell motility, reproductive

CC and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to

CC CYSKP-16 can be used to treat or prevent disorders associated with

CC decreased expression or activity of CYSKP (claimed), for example,

CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,

CC cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,

CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,

CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma

CC and trauma. CYSKP antagonists can be used to treat or prevent a

CC disorder associated with increased expression or activity of CYSKP

CC (claimed).

XX Sequence 1005 AA;

SQ

Query Match 35.3%; Score 689.5; DB 21; Length 1005;

Best Local Similarity 42.8%; Pred. No. 3.2e-61;

Matches 148; Conservative 58; Mismatches 129; Indels 11; Gaps 6;

QY 34 TPGSKLVSTIKIOMLDDTQEAPEVQAPGKVLDDAVCNHLNLVEGDYFGLFPDHHKITV 93

DB 212 tkktktvqcvktllldgteyscdlekhakgqvlfdkvcchlnllekdylfqlqespeqkn 271

QY 94 WLDLLLPVVKQIRPKHVVKVFPVPPDHTQLOBELTRYLFALQVKODLAQGLTCLND 153

DB 272 wldpakeikrqlnlpwft-fnvklypppsqlteditryfcicqlrqdiasgrlpcsf 330

QY 154 TSAALLISHIVQSEIGDFDEALDREH---LAKNKYIP-OODALEDKIVFHHNHIGOTP 208

DB 331 vthallgsytlqaelgoyd---peehgsidlsefqaftqtkeekvaalhhthrglsp 387

QY 209 AESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQFTKINAFNNAKVRK 268

DB 388 aqadsqflenakrlsmgyvdlhhakdsegvdklgyvcangllykdrlnrnfawpklk 447

QY 269 LSPKRRFLIKLRPDANSAYQDTLEFLMASRDECKSFWKICVEHHAFFRLEFEPKPKPKP 328

DB 448 isykrnfyikvrpaelqfestigfklnpuraakrlwkvcehhtfyrlv-speqppka 506

QY 329 VLFSGSSFRFSRGRTQKQVLDYVKEGGHKKVQPERKHSKIHSIRSL 374

DB 507 kftlgskfrysgtqaqrqastlldrpaphfertsskrvs-rsl 551

RESULT 4

ID AAB95258

XX AAB95258 standard; Protein; 687 AA.

AC AAB95258;

XX 26-JUN-2001 (first entry)

DT Human protein sequence SEQ ID NO:17435.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

OS EP1074617-A2.

PN 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.


```
XX Disclosure; Fig 1A-B; 12pp; English.
PS
CC A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the
CC dephosphorylation of proteins in which tyrosyl residues have been
CC phosphorylated through the action of a protein tyrosine kinase
CC (PTK). It localises to focal adhesions, a major site of action of
CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)
CC obt'd. from HeLa cells. The PTPH1 cDNA can be incorporated into a
CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian
CC in sufficient quantities to overcome or counteract PTK activity.
CC Phosphorylation of tyrosine residues at abnormal levels is
CC prevented or reversed, resulting in the prevention or reversal of
CC malignancy of cells.
XX
SQ Sequence 913 AA;

Query Match 25.9%; Score 506.5; DB 18; Length 913;
Best Local Similarity 35.0%; Pred. No. 1.4e-42;
Matches 119; Conservative 68; Mismatches 134; Indels 19; Gaps 9;

QY 27 RGQKPPPTPSGKLVSIKIQMLDDTQEAFFVPPQAPGKVLDDAVCNHNLNVEGDYFGLEFP 86
Db 16 rtseipkektseivcsihfidgvvqtkvtdqgtgvlldmvmhnhlgvtekeyfglqhd 75
QY 87 DHKKITV-WLDLLKPIVKQIRRPKHVVVVFVVFPPDHTQQLBELTRYLFALQVKODLA 145
Db 76 ddsdvsprwleaskpikrkqkgfpcctlhfrvrfipdpntlqqeqtrhlyflqlkmdic 135
QY 146 QGRITCNDTSAALLSHIVQSEIGDFDEALDRE-HLAKNKYIPOQDALED---KIVEFHH 201
Db 136 egritcplnsavvlasyavqshgfydgnssihhpgyilsdshfipdqn--edfltkvesl 193
QY 202 NHIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQTKINAF 261
Db 194 qhsglkqseaseacyiniartldfygvelhsgdrldhldlmigiasagvavrykictsfy 253
QY 262 NNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSPWKVCVEHAFRLEFEE 321
Db 194 qhsglkqseaseacyiniartldfygvelhsgdrldhldlmigiasagvavrykictsfy 253

RESULT 6
AAY25156
ID AAY25156 standard; Protein; 913 AA.
XX
AC AAY25156;
XX
XX
DT 07-SEP-1999 (first entry)
XX
DE Human PTPH1 protein.
XX
KW PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;
KW localisation; treatment; overexpression; oncogenic; cell transformation;
KW prevention; phosphotyrosine; disease; malignant.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Region 30..357
FT /note= "region of homology to the N-terminal domain
FT of band 4.1, ezrin and talin. This region is
FT known to be important for localisation to focal
FT adhesions"
FT
FT Modified-site 372
FT /note= "potential p34cdc2 phosphorylation site"
FT
FT Modified-site 381
FT /note= "potential p34cdc2 phosphorylation site"
FT
```

```
FT Modified-site 424..428
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 434
FT /note= "potential p34cdc2 phosphorylation site"
FT Modified-site 438..442
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 489..492
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 514..518
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 543..547
FT /note= "potential casein kinase II phosphorylation site"
FT Modified-site 607..610
FT /note= "potential casein kinase II phosphorylation site"
XX
XX US5863781-A.
XX
XX 26-JAN-1999.
XX
XX 04-DEC-1996; 96US-0759536.
XX
XX 01-MAR-1991; 91US-0663579.
XX 14-MAR-1990; 90US-0494036.
XX 16-AUG-1993; 93US-0107420.
XX 04-DEC-1996; 96US-0759536.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX
XX Tonks NK;
XX
XX WPI; 1999-131308/11.
XX N-PSDB; AAX78463.
XX
XX Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells
XX
XX Claim 1; Fig 1A-B; 12pp; English.
XX
XX This sequence represents a novel protein tyrosine phosphatase, PTPH1,
XX isolated from HeLa cells. The protein of the invention appears to
XX localise to focal adhesions and is therefore potentially useful in the
XX treatment of cancer. Overexpression of PTPH1 can be used to counter the
XX effects of oncogenic protein tyrosine kinases such as those of
XX transforming viruses and for interfering with or reversing cell
XX transformation. This would provide a means of preventing or reversing
XX abnormally high levels of phosphotyrosine associated with any disease or
XX condition such as preventing or reversing malignancy associated with the
XX activity of a protein tyrosine kinase.
XX
XX Sequence 913 AA;

Query Match 25.9%; Score 506.5; DB 20; Length 913;
Best Local Similarity 35.0%; Pred. No. 1.4e-42;
Matches 119; Conservative 68; Mismatches 134; Indels 19; Gaps 9;

QY 27 RGQKPPPTPSGKLVSIKIQMLDDTQEAFFVPPQAPGKVLDDAVCNHNLNVEGDYFGLEFP 86
Db 16 rtseipkektseivcsihfidgvvqtkvtdqgtgvlldmvmhnhlgvtekeyfglqhd 75
QY 87 DHKKITV-WLDLLKPIVKQIRRPKHVVVVFVVFPPDHTQQLBELTRYLFALQVKODLA 145
Db 76 ddsdvsprwleaskpikrkqkgfpcctlhfrvrfipdpntlqqeqtrhlyflqlkmdic 135
QY 146 QGRITCNDTSAALLSHIVQSEIGDFDEALDRE-HLAKNKYIPOQDALED---KIVEFHH 201
Db 136 egritcplnsavvlasyavqshgfydgnssihhpgyilsdshfipdqn--edfltkvesl 193
QY 202 NHIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQTKINAF 261
Db 194 qhsglkqseaseacyiniartldfygvelhsgdrldhldlmigiasagvavrykictsfy 253
QY 262 NNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSPWKVCVEHAFRLEFEE 321
Db 194 qhsglkqseaseacyiniartldfygvelhsgdrldhldlmigiasagvavrykictsfy 253
```


CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

XX
 SQ Sequence 635 AA;

Query Match 18.4%; Score 359; DB 21; Length 635;
 Best Local Similarity 27.9%; Pred. No. 9.9e-28;
 Matches 107; Conservative 71; Mismatches 148; Indels 58; Gaps 13;

QY 7 RPTPGSRLGAPEN-----GISTLERGQPPPTPSGLVSIKIQMLDDTQAEFVPPQR 59
 Db 19 rstagaglgqlrccwivvfgiedthqkpkmp-----kpinrvrtmdaele-faiqp 72

QY 60 APGVLDAYCNHNLNVEGDFYGLFPPDHKKITWLDLLKPI-VKQIRPKHVHVVFVK 118
 Db 73 ttgqlfdqvvtiglrrevvyfghyvdnkgfptwklkdkksaqevrkenplqfkfrak 132

QY 119 FFPDP-HTOLQBELTRVLFALQVQDIAQGLTCNDTSAALLISHIVQSEIGDFDEALDR 177
 Db 133 fypedvaeelqgdtqklflqkqegilsdeiyppetavllgsyavqakfgdynkevkh 192

QY 178 E-HLAKNKYIPQ-----QDALEDKIVEFHNNHIGQTPAESDFOLLEIARRLEMYGI 227
 Db 193 sgyllserlipqvmqdkhitrqwedriqwhaehrgmlkdnamleylkiadqlenygi 252

QY 228 RLHPAKDREGTKINLAVANTGIIVFGFTKIN---AFNNAKVRKLSFKRRFLI----KL 280
 Db 253 nyfeiknkgtdlwlvgdaigltnyekddkltpkigfpwseirnisfndkfkvpkldkk 312

QY 281 RPDANSAYQDTLEFLMASRDFCKSFWKICVVEHHAFFELFEEPKPKPKPVLFSGSFRFS 340
 Db 313 apd-----fvyaprlrlnkrlqlcmgnhely-mrrrkpdktiev----- 351

QY 341 GRTQKQVLDYVKEGGHKKVQFERK 364
 Db 352 ---qgmkaqaareekhqk-qlerq 370

RESULT 9
 AAW67438
 ID AAW67438 standard; Protein; 1174 AA.
 XX
 AC AAW67438;
 XX
 DT 02-MAR-1999 (first entry)
 DE
 DE Human protein tyrosine phosphatase D1 protein.
 XX
 KW Human; protein tyrosine phosphatase; skeletal muscle; hybridisation;
 KW glycoprotein; probe.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 24..339
 FT Binding-site /note="ezrin-like domain"
 FT 158..161
 FT Binding-site /note="SH2 binding motif"
 FT 207..210
 FT Binding-site /note="SH2 binding motif"
 FT 217..220
 FT Binding-site /note="SH2 binding motif"
 FT 336..338
 FT Modified-site /note="PYX motif, putative site of Tyr phosphorylation"
 FT 354..356
 FT Modified-site /note="PYX motif, putative site of Tyr phosphorylation"
 FT 408..410
 FT Modified-site /note="PYX motif, putative site of Tyr phosphorylation"
 FT 523..525

FT Modified-site /note="PYX motif, putative site of Tyr phosphorylation"
 FT 525..527
 FT Binding-site /note="PYX motif, putative site of Tyr phosphorylation"
 FT 565..574
 FT Modified-site /note="SH3 binding domain"
 FT 568..570
 FT Region /note="PYX motif, putative site of Tyr phosphorylation"
 FT 712..722
 FT Domain /note="acidic stretch"
 FT 921..1166
 FT /note="protein tyrosine phosphatase domain"
 XX
 PN US5831009-A.
 XX 03-NOV-1998.
 XX 22-MAY-1995; 95US-0446345.
 PF 27-APR-1994; 94US-0234440.
 PR 15-AUG-1992; 92US-0923740.
 PR 22-MAY-1995; 95US-0446345.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PI Moller KB, Moller NPH, Ullrich A;
 XX
 WPI: 1998-609316/51.
 DR N-PSDB; AAV34368.
 XX
 PT Protein tyrosine phosphatase D1 - useful in screening assays for
 PT e.g. agonists or antagonists
 XX
 PS Claim 1; Fig 5A-J; 61pp; English.
 CC
 CC This is the sequence of the complete human protein tyrosine phosphatase
 CC D1 (PTP-D1). The coding sequence was isolated from a human skeletal
 CC muscle cDNA library. The phosphatase is used e.g. in a hybridisation
 CC assay for detecting a nucleic acid encoding a normal or mutant PTP-D
 CC protein or glycoprotein in a cell or subject, using a nucleic acid
 CC comprising at least part of a normal or mutant PTP-D protein or
 CC glycoprotein as a probe; or a screening assay for a compound capable
 CC of binding to a PTP-D protein or glycoprotein, using a PTP-D protein
 CC or glycoprotein or the compound-binding portion of it attached to a
 CC solid support.
 XX
 SQ Sequence 1174 AA;

Query Match 18.2%; Score 356; DB 19; Length 1174;
 Best Local Similarity 27.1%; Pred. No. 5e-27;
 Matches 94; Conservative 70; Mismatches 147; Indels 36; Gaps 8;

QY 25 LERGQPPPTPSKIVSIKIQMLDDTQAEFVPPQRAPGVLLDAVCNHLNVEGDFYGLE 84
 Db 9 lkrtrrytvssksclva-riqlinnfevftsvestqgesleavagrieirevtyfslw 67

QY 85 FPDHKKITVWLDLLKPIVKQI-RRPKHVHVVFVPPDPHTQOELRYLFALQVQD 143
 Db 68 yynkqqrtrwvdeklpkqlkqdyaleptvyfgyvsvsqqlqceitryqylqlkdd 127

QY 144 LAQGLTCNDTSAALLISHIVQSEIGDFDEALDREHLAKNKYIP-----QDALED---X 195
 Db 128 ilegspictleqaiqlaglavqadfgdgyesqdfqlkfalfpvgwlgdekveeatqk 187

QY 196 IVEFHNNHIGQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIIV-FQ 254
 Db 188 vallhdkyrgltapdaemlymgevermdgyeesypaksgqsdgisgaclegifvkhkn 247

QY 255 FTKINAFNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVVEHHA 314
 Db 248 grhpvfvfrwhdianmshnksffalel---ank---estiqfqtmetakylwrlcvarkh 302

QY 315 FFLRFE-----EPKPKPKPVLFSGSFRFSG 341


```

Db 62 dkksaqevrkenplqgfkfrakfyedvaeellqditqklflqvkegilsdeiyccpet 121
QY 156 AALLISHIVOSEIGDFDEALDRE-HLAKNKYIPQ-----ODALEDKIVEFHNNHIG 205
Db 122 avlgsyavqakfgdyknkngyslsserlfpqrvmqghkltrdqwedrlqvwahaeirg 181
QY 206 QTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILVFGQFTKIN---AFN 262
Db 182 mlkdnamleykiaqdlmgyinyfeiknkgtdlwgvdalgnlyekdkktpkigfp 241
QY 263 WAKVKLSFRKRFLI---KLAPDANSAYDTLEFLMASRDFCKSWFKTCVEHHAFRL 318
Db 242 wseirnisfndkfkvikpidkaped-----fvyaprlrlnkrilqicmgnhely-- 291
QY 319 FEERPKPKPVLSRGSSFRSGRTOKQVLDYVKEGGHKVQPERK 364
Db 292 mrrrkpdrtiev-----qmqkaqaareekhqq-qierq 321

RESULT 12
AAW94458
ID AAW94458 standard; Protein; 590 AA.
XX AC AAW94458;
XX DT 16-APR-1999 (first entry)
XX DE Human neurofibromatosis type 2 transcript isoform II protein.
XX KW Human; neurofibromatosis type 2; NF2; tumour suppressor; cancer;
XX KW diagnosis; gene therapy.
XX OS Homo sapiens.
XX FT Key Location/Qualifiers
FT Misc-difference 570..571
FT /note= "in the Sequence Listing an additional Pro is
FT between the Gly and Ser residues given in
FT Figure 7"
XX US5872214-A.
XX PD 16-FEB-1999.
XX PF 04-APR-1996; 96US-0628145.
XX PR 10-JAN-1994; 94US-0179738.
XX PR 04-APR-1996; 96US-0628145.
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX PI Bianchi AB, Kley NA, Seizinger BR;
XX WPI; 1999-166715/14.
XX DR N-PSDB; AAX04296.
XX PT Proteins from neurofibromatosis type 2 transcript isoforms - used
XX for diagnosis or inhibition of tumours, and generation of antibodies
XX Claim 4; Fig 7; 45pp; English.
XX The present sequence represents human neurofibromatosis type 2 (NF2)
XX transcript isoform II. NF2 polynucleotides can be used for diagnosing
XX NF2 diseases, for inhibiting growth of tumours associated with NF2
XX mutations (including expression from cDNA introduced in gene therapy
XX vectors) and to raise antibodies (useful as tumour targeting agents,
XX since specific isoforms are often tumour-specific) and as immunoassay
XX reagents for detecting NF2-expression products. NF2 is a tumour
XX suppressor protein, and so has anticancer activity.
XX Sequence 590 AA;

```

```

Query Match 16.8%; Score 328; DB 20; Length 590;
Best Local Similarity 27.2%; Pred. No. 1.3e-24;
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGOKPPPTSGKLVYSIKIOMLDDTQBAFEVFPQAPQKVLDDAVCNHLNLVEGDFG 82
Db 12 ssikr--kqpkttftrivtm-----daemefncemkwgkdlfdlvcrtltgretwffg 63
QY 83 LEFPDHHKITVWLDLKLPIV-KOIRRPKHVVVVKVVFPPD-HTOLQEEELTRYLFALQV 140
Db 64 lqy-tlkdtvawlkmdkklvhdvskeepvtfhflakfypenaeeelvgeltqnlflqv 122
QY 141 KDLAOGRLTCNDTSALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190
Db 123 kqildkciyccpeasvllasyavakgydpsvnhrgflaqeellpkrvlnyqmtpe 182
QY 191 ALEDKIVEFHNNHIGQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 mweiritawyaehrgrardeameyikiaqdlmgyinyfairnkkgtelllgvdalgh 242
QY 251 VFOGFTKIN---AFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSWFK 307
Db 243 iydpentrlpkisfpwneirnisysdkeftik----pldkkidvfkfnssklrvnkllq 298
QY 308 ICVEHHHAFF 316
Db 299 lcignhldf 307

RESULT 13
AAW09648
ID AAW09648 standard; Protein; 591 AA.
XX AC AAW09648;
XX DT 14-MAY-1997 (first entry)
XX DE Human merlin protein isoform II, a tumour suppressor.
XX KW NF2; neurofibromatosis type 2; multiple tumours; nervous system;
XX KW bilateral vestibular schwannoma; acoustic neuroma; cranial nerve;
XX KW meningioma; lens opacity; chromosome region 22q12; tumour suppressor;
XX KW merlin; moesin-radixin like protein; alternative splicing;
XX KW diagnosis; cancer; neoplasia; autosomal; dominant; hereditary.
XX OS Homo sapiens.
XX XX US5578462-A.
XX PD 26-NOV-1996.
XX PF 10-JAN-1994; 94US-0179738.
XX PR 10-JAN-1994; 94US-0179738.
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX PI Bianchi AB, Kley NA, Seizinger BR;
XX WPI; 1997-020406/02.
XX DR N-PSDB; AAT47832.
XX PT New isolated mouse and human NF2 transcript isoforms - used to
XX develop prods. for the diagnosis and treatment of neurofibromatosis
XX type 2 diseases.
XX PS Claim 4; Column 45-48; 46pp; English.
XX CC AAW09648 is the merlin protein product of human NF2 (neurofibromatosis
XX type 2) gene isoform II. NF2 is an autosomal, dominantly inherited
XX disorder characterised by multiple tumours of the central nervous
XX system, predominantly bilateral vestibular schwannomas (acoustic

```

CC neuromas) of the eighth cranial nerve. Other symptoms of NF2 include
 CC cranial meningiomas, spinal nerve root schwannomas and presenile lens
 CC opacities. The NF2 gene, mapped to chromosomal region 22q12 between the
 CC loci D22S1 and D22S28, acts as a tumour suppressor. The NF2 gene is
 CC alternatively spliced resulting in three different isoforms encoding
 CC three different proteins, merlin isoforms I-III, which are likely to
 CC have differing functions. Merlin stands for moesin-erzin-radixin like
 CC protein, so called due to substantial homology with these three
 CC proteins. NF2 isoform II cDNA and protein are useful in diagnosing NF2
 CC disease. Merlin protein products act as tumour suppressors and can be
 CC used to suppress tumour growth, as can the cDNA sequence in gene therapy
 CC applications. Antibodies raised against merlin proteins are useful as
 CC tumour targeting agents. The human NF2 isoform was identified in a
 CC foetal mouse brain cDNA library.

XX Sequence 591 AA;

Query Match 16.8%; Score 328; DB 18; Length 591;

Best Local Similarity 27.2%; Pred. No. 1.3e-24;

Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGQPPPTPSGKLVSIKIQMLDDTQEAPEVFPQAPGKVLDDAVCNHNLNVEGDYFG 82

DB 12 ssLkr--kqpkftvriutm-----daemefncemkkgkldfvlvrtlgretwffg 63

QY 83 LEFPDHKKITVWLDLLKPIV-KQIRPKHVVKVVFVFPDPD-HTQLOEELTRYLFALQV 140

DB 64 lqy-tikdtvawlkmdkvdhdskeepvtfhflakypenaeeelvgelqthifflqv 122

QY 141 KQDLAQGRITCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190

DB 123 kkgildelkycpeasvllasyavqakgydpsvhhkgflaqeelpkrvinlygmtpe 182

QY 191 ALEDKIVEFHNNHIGQTPAESDFQLLETARLEMGIRLHPAKDREGTKINLAVANTGTL 250

DB 183 mweeritawyaehrgrardeaemeylkiaqdlmgyvnyfairnkkgtellgvdalgh 242

QY 251 VFQGTFTKN---AFNWKVKRLSEKRRFLIKLRPDANSAYODTLEFLMASRDFCKSPWK 307

DB 243 lydpentpklspwneirnisysdkeftik----pidkklidvfkfnsklrvnklliq 298

QY 308 ICVEHHAF 316

DB 299 lcighndlf 307

RESULT 14

ID AAR60398 standard; Protein; 595 AA.

AC AAR60398;

20-APR-1995 (first entry)

DE Merlin protein encoded by cDNA clone JJR-1.

KW Polymerase chain reaction; PCR; amplify; primer; bi-lateral schwannoma;

KW sequence-tagged site assay; chromosome 22; NF2; deletion; hearing loss;

KW neurofibromatosis; merlin; moesin-erzin-radixin-like protein; D22S28;

KW tumour suppressor; activity; meningioma; cytoskeleton; gene therapy;

KW merlin-associated tumour; D22S1; posterior capsular lens opacity;

KW deafness; balance disorder; paralysis; ss.

XX Homo sapiens.

OS EP613945-A.

PN 07-SEP-1994.

XX 25-FEB-1994; 94EP-0301367.

XX 25-FEB-1993; 93US-0022034.

PR 04-MAR-1993; 93US-0026063.

PR 19-AUG-1993; 93US-0108808.

XX 22-DEC-1993; 93US-0171718.

PA (GCHO) GEN HOSPITAL CORP.

XX Gusella JF, MacCollin MM, Trofatter JA;

XX WPI; 1994-272992/34.

DR N-PSDB; AAQ71077.

XX The tumour suppressor gene merlin - for treatment and diagnosis

PT of tumours and neurofibromatosis (NF2)

XX Example 3; Page 35-38; 86pp; English.

XX This sequence is encoded by the clone JJR-1 and represents the merlin

CC protein. The cDNA sequence contains eight cloned exons sequences and

CC encodes this protein of 69 kD. The cDNA contains no evidence of a

CC poly-A tail and spans at least 50 kb of genomic DNA. NF2 is a neuro-

CC fibromatosis which is characterised by bi-lateral schwannomas. The

CC NF2 "gene" has been shown by linkage studies to be assigned to

CC chromosome 22. The missing or mutated gene in NF2 patients has been

CC shown to be the merlin gene. The gene encodes a protein, merlin

CC (moesin-erzin-radixin-like protein), which possesses tumour suppressor

CC activity, and whose tumour suppressor activity is mediated by inter-

CC actions with the cytoskeleton. The merlin gene is found on chromosome

CC 22 between the known markers D22S1 and D22S28. In patients suffering

CC from NF2, the merlin gene is either lost or mutated. A mutant merlin

CC protein may be encoded by a gene in which a mutation of A to T at the

CC first position of the codon encoding amino acid 220 causes the

CC substitution of Tyr for Asn. The merlin gene may be used in gene

CC therapy for the treatment of a merlin-associated tumour or NF2, or

CC for prevention of schwannoma, meningioma, posterior capsular lens

CC opacities, deafness or hearing loss, balance disorders or paralysis.

XX Sequence 595 AA;

Query Match 16.8%; Score 328; DB 15; Length 595;

Best Local Similarity 27.2%; Pred. No. 1.3e-24;

Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGQPPPTPSGKLVSIKIQMLDDTQEAPEVFPQAPGKVLDDAVCNHNLNVEGDYFG 82

DB 12 ssLkr--kqpkftvriutm-----daemefncemkkgkldfvlvrtlgretwffg 63

QY 83 LEFPDHKKITVWLDLLKPIV-KQIRPKHVVKVVFVFPDPD-HTQLOEELTRYLFALQV 140

DB 64 lqy-tikdtvawlkmdkvdhdskeepvtfhflakypenaeeelvgelqthifflqv 122

QY 141 KQDLAQGRITCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190

DB 123 kkgildelkycpeasvllasyavqakgydpsvhhkgflaqeelpkrvinlygmtpe 182

QY 191 ALEDKIVEFHNNHIGQTPAESDFQLLETARLEMGIRLHPAKDREGTKINLAVANTGTL 250

DB 183 mweeritawyaehrgrardeaemeylkiaqdlmgyvnyfairnkkgtellgvdalgh 242

QY 251 VFQGTFTKN---AFNWKVKRLSEKRRFLIKLRPDANSAYODTLEFLMASRDFCKSPWK 307

DB 243 lydpentpklspwneirnisysdkeftik----pidkklidvfkfnsklrvnklliq 298

QY 308 ICVEHHAF 316

DB 299 lcighndlf 307

RESULT 15

ID AAW94459 standard; Protein; 596 AA.

XX AAW94459;

Job time: 334 sec

```
XX 16-APR-1999 (first entry)
XX Human neurofibromatosis type 2 transcript isoform I protein.
XX
XX Human; neurofibromatosis type 2; NF2; tumour suppressor; cancer;
XX diagnosis; gene therapy.
XX
XX Homo sapiens.
XX
XX US5872214-A.
XX
XX 16-FEB-1999.
XX
XX 04-APR-1996; 96US-0628145.
XX
XX 10-JAN-1994; 94US-0179738.
XX
XX 04-APR-1996; 96US-0628145.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Bianchi AB, Kley NA, Seizinger BR;
XX
XX WPI; 1999-166715/14.
XX
XX
XX Proteins from neurofibromatosis type 2 transcript isoforms - used
XX for diagnosis or inhibition of tumours, and generation of antibodies
XX
XX Example 1; Fig 1; 45pp; English.
XX
XX The present sequence represents human neurofibromatosis type 2 (NF2)
XX transcript isoform I. NF2 polynucleotides can be used for diagnosing
XX NF2 diseases, for inhibiting growth of tumours associated with NF2
XX mutations (including expression from cDNA introduced in gene therapy
XX vectors) and to raise antibodies (useful as tumour targeting agents,
XX since specific isoforms are often tumour-specific) and as immunoassay
XX reagents for detecting NF2-expression products. NF2 is a tumour
XX suppressor protein, and so has anticancer activity.
XX
XX Sequence 596 AA;

Query Match 16.8%; Score 328; DB 20; Length 596;
Best Local Similarity 27.2%; Pred. No. 1.3e-24;
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERQKPPPTPSGKLVSIKQLMLDDTQEAPEVQAPAGKVLDDAVCNHLNVEGDYFG 82
DB 12 ssikr--kqpkftftrivlm-----daemefncemkwgkdlfdlvcrtlgretwffg 63
QY 83 LEFPDHKKITVWLDLKPIV-KQIRPKHVWVVKVFPDPD-HTOLOBELTRYLFALOV 140
DB 64 lgy-tikdtvawlkmdkdvldhdvskeepvtfhflakfypenaeelqvqitqhlfflgy 122
QY 141 KDLAQGRITCNDTSAAALISHIVQSEIGDFEAL-DREHLAKNKYIPOQ-----D 190
DB 123 kqildeklycypeasvllasyavqakgydpsvvhkgfqlaqeellpkrvlnlyqmtpe 182
QY 191 ALEDKIVEFHNNHIGTPAESFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 183 mweeritavyaehrgardeaemeykiaqdiemygvnyfairnkkgtellgvdalgh 242
QY 251 VFQGFTKIN--AFNNAKVKRLSFKRFLIKLRPDANSAYODTLEFLMASRDFCKSWK 307
DB 243 iydpenrltpklsfpwneirnisysdkeftik----pldkkidvfkfnsskkrvnklliq 298
QY 308 ICVEHHAFF 316
DB 299 lcignhldf 307
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:55:08 ; Search time 36.72 Seconds
(without alignments)
229.200 Million cell updates/sec

Title: US-09-555-342A-2_COPY_1_374

Perfect score: 1956

Sequence: 1 MGIEQPTFGSLGAPENS.....GHKKVQFERKHKHSIRSL 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	646	33.0	345	2	US-08-446-345-40
2	612	31.3	288	3	US-09-100-804-14
3	522	26.7	247	3	US-09-100-804-13
4	500.5	25.6	288	3	US-09-100-804-15
5	499	25.5	342	2	US-08-446-345-39
6	356	18.2	1174	2	US-08-446-345-36
7	344	17.6	296	3	US-09-100-804-12
8	328	16.8	591	1	US-08-179-738-10
9	328	16.8	591	2	US-08-628-145-10
10	328	16.8	595	1	US-08-171-718-16
11	328	16.8	595	3	US-08-478-087-16
12	328	16.8	596	1	US-08-179-738-3
13	328	16.8	596	2	US-08-628-145-3
14	325	16.6	584	1	US-08-179-738-7
15	325	16.6	584	2	US-08-628-145-7
16	323	16.5	591	1	US-08-179-738-5
17	323	16.5	591	2	US-08-628-145-5
18	323	16.5	596	1	US-08-179-738-2
19	323	16.5	596	2	US-08-628-145-2
20	265	13.5	303	3	US-09-100-804-11
21	265	13.5	2465	3	US-08-596-291-3
22	265	13.5	2465	3	US-09-100-804-3
23	265	13.5	2466	3	US-09-080-855-12
24	265	13.5	2466	5	PCT-US94-09943-2
25	265	13.5	2485	4	US-09-230-640-46
26	91	4.7	522	1	US-08-625-322-2
27	89.5	4.6	212	4	US-09-259-109-2

28	87	4.4	1164	4	US-08-923-992A-10	Sequence 10, Appl
29	86	4.4	472	1	US-08-203-905B-14	Sequence 14, Appl
30	85.5	4.4	1704	4	US-08-485-355B-40	Sequence 40, Appl
31	84.5	4.3	1462	3	US-07-792-600-31	Sequence 31, Appl
32	84.5	4.3	1462	3	US-09-157-021-31	Sequence 31, Appl
33	84.5	4.3	1462	3	US-09-156-842-31	Sequence 31, Appl
34	84	4.3	984	1	US-08-242-932-2	Sequence 2, Appl
35	84	4.3	984	1	US-08-714-481-2	Sequence 2, Appl
36	84	4.3	984	5	PCT-US95-06111-2	Sequence 2, Appl
37	84	4.3	1164	4	US-08-923-992A-2	Sequence 2, Appl
38	83.5	4.3	633	2	US-08-736-770-3	Sequence 3, Appl
39	82	4.2	1128	4	US-08-923-992A-6	Sequence 6, Appl
40	82	4.2	1513	5	PCT-US93-03076-2	Sequence 2, Appl
41	81.5	4.2	517	2	US-08-967-508-19	Sequence 19, Appl
42	81.5	4.2	517	3	US-08-967-506-19	Sequence 19, Appl
43	81.5	4.2	517	5	PCT-US94-02552-19	Sequence 19, Appl
44	81.5	4.2	559	2	US-08-967-508-9	Sequence 9, Appl
45	81.5	4.2	559	3	US-08-967-506-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-446-345-40
; Sequence 40, Application US/08446345
; Patent No. 5831009
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASES PTP-D1
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-446-345-40

Query Match 33.0%; Score 646; DB 2; Length 345;
Best Local Similarity 40.9%; Pred. No. 4.5e-59;

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,345
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,440
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 7683-054
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-446-345-39

Query Match 25.5%; Score 499; DB 2; Length 342;

Best Local Similarity 34.7%; Pred. No. 9,le-44;

Matches 114; Conservative 66; Mismatches 133; Indels 16; Gaps 7;

QY 27 RGQKPPPTSGKLVSTIKIQLDDTQFAFVPPQAPGKVLDDAVCNHNLNVEGDYGLFPP 86
DB 16 RISELPEKTRSEVICSIFHVGQVOTFKVTKQDTGQVLDWMVHNLGVTKEYFGLQHD 75
QY 87 DHKKITV-WLDDLKPIVKQIRRPKHVVVKKVFFPPDPDHTLQEEELRYLFAIQVKQDILA 145
DB 76 DSDVSPRLEASLPKQKGGFCTLHFRVRFPIPDNTLQEQTRHLYLQLKMDIC 135
QY 146 QGRLCNTDTSALLSHIVQSIGDFEALDRE-HLAKNKYIPQDALED---KIVEFH 201
DB 136 EGRLTCLPLNSAVLASYAFHGFDYNSIHHPGYSLSHSHFIPDQ--EDFLTQVESLHE 193
QY 202 NHIGQTPAESDFOLLEIARRLEMVIGRLHPAKDREGTKINLAVANTGILVFOGFTKINAF 261
DB 194 QHSGLKQSAESCYNIARTLDFYGVLSGRDLNLDLMIGIASAGVAYRYKICTSY 253
QY 262 NNAVKRLSFKRRRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFTLPEE 321
DB 254 PWNILKISFKRKKFTIHQKQAESREHIVAFNMLNYSRCKNLWKSCVEHHTFQA-KK 312
QY 322 PKPKPKPVL---FSRGSFRFSGRQKQV 347
DB 313 LLPQEKVLSQYWTGSR-----NTKKS 336

RESULT 6

US-08-446-345-36

; Sequence 36, Application US/08446345

; Patent No. 5831009

; GENERAL INFORMATION:

; APPLICANT: Ullrich, Axel

; APPLICANT: Moller, Niels P.H.

; APPLICANT: Moller, Karin B.

; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE

; TITLE OF INVENTION: PHOSPHATASES PTP-DI

; NUMBER OF SEQUENCES: 41

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1174 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-446-345-36

Query Match 18.2%; Score 356; DB 2; Length 1174;

Best Local Similarity 27.1%; Pred. No. 4.7e-28;

Matches 94; Conservative 70; Mismatches 147; Indels 36; Gaps 8;

QY 25 LERGQPPPTSGKLVSTIKIQLDDTQFAFVPPQAPGKVLDDAVCNHNLNVEGDYGLF 84
DB 9 LKTRRYTVSSKCLVA-RIQLLNNEFVEFTLSVESTGQESLEAQAORLELREVTYFSLW 67
QY 85 FPDHKKITVWLDLKPVKQI-RRPKHVVVVKKVFFPPDPDHTLQEEELRYLFAIQVKOD 143
DB 68 YYNQQRNRWDLERPLKKQLDKYALEPTVYGVVFPVPSQLQOEITRYQYVQLKKD 127
QY 144 LAQGRLTCDNTSAALLSHIVQSIGDFEALDREHLAKNKYIP-----QDALED---K 195
DB 128 ILEGIPCTLEQAIQAGLAVQDFGDFQYESQDFLQKFALEPVGWLDQEKVLEATQK 187
QY 196 IVEFHNNHIGQTPAESDFOLLEIARRLEMVIGRLHPAKDREGTKINLAVANTGILV-FQ 254
DB 188 VALLHQYRGLTPADAEMLYMQEVRMDGYGESYPAKDSQSDISIGACLEGIFVKHN 247
QY 255 FTKINAFNWKVRKLSFKRRRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHA 314
DB 248 GRHPVYFRWHDIANMHNKSFFALEL---ANK--EETIQFQTEDEMTAKYIWRLCVARHK 302
QY 315 FPLRFE-----EPKPKPKVLFSGSSFRFSG 341
DB 303 FYRLNQCNLQOTVTVNPIRRSSSRMSLPPKQPYVMPPPPQLHYNG 349

RESULT 7

US-09-100-804-12

; Sequence 12, Application US/09100804

; Patent No. 6066472

; GENERAL INFORMATION:

; APPLICANT: GONEZ, LEONEL JORGE

; APPLICANT: SARAS, JAN

; APPLICANT: CLAESSON-WELSH, LENA

; APPLICANT: HELDIN, CARL-HENRIK

; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN


```
Db 299 LCIGNHDLF 307
RESULT 9
US-08-628-145-10
; Sequence 10, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
;
US-08-628-145-10
Query Match 16.8%; Score 328; DB 2; Length 591;
Best Local Similarity 27.2%; Pred. No. 1.4e-25;
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGQKPPPTSGKLVSIKIQMLDDTQAEFVPPORAPGKVLDDAVCNHLNLVEGDYFG 82
Db 12 SSLKR--KPKFTTFRIVTM-----DAEMEFNCENMKWKGKDLFDLVCRTLGLRETWFFG 63

QY 83 LEPPDHKKTTVWLDLKKPIV-KQIRPKHVYKVVVKKFFPPD-HTQLQBELTRYLFALQV 140
Db 64 LQY-TIKDTVAWLKMDKKVLDHDVSKKEEPTVTHFLAKFPENAEELVQEIQHLEFFLQV 122

QY 141 KDLAQRGRTCNDSALLISHIVQSEIGDFEAL--DREHLAKNKYIPQO-----D 190
Db 123 KQILDEKICYPEASVLAASAVQAKYGDYDFSVHKRGLAQEELTPKRVINLYQMTPE 182

QY 191 ALEDKIVFHHNHIGTQPAESDFQLLEIARRLEMYGIRLHPAKDRGKTINLAVANTGIL 250
Db 183 MWEERITAVAEHRGRARDEAEEMYLKIAQDLEMYGVNYFAIRNKKGTETLLGVDALGLH 242

QY 251 VFOGFTKIN---AFNWAQVKLSFKKRFLIKLRPDANSAYQDTLEFLMASRDFCSFWK 307
Db 259 LCIGNHDLF 307
```

Db 123 KKQILDEKIYCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEBELLKRVINLYQMTPE 182
Qy 191 ALEDKIVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERITAWAEHRGRARDEAEMEYLKIAQDLEMYGVNYFAIRNKKGTTELLIGVDALGLH 242
Qy 251 VFOGFTKIN--AFNNAKVRKLSFKRKRFLIKLRDPDANSAYQDTLEFLMASRDFCKSEWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSKDEFTIK----PLDKIDVFKFNSSKLRVKNKLIQ 298
Qy 308 ICVEHHAFF 316
Db 299 LCIGNHDLF 307
RESULT 11
US-08-478-087-16
; Sequence 16, Application US/08478087
; Patent No. 6077685
; GENERAL INFORMATION:
; APPLICANT: Trofatter, James A.
; APPLICANT: MacCollin, Mia M.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,087
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,718
; FILING DATE: 22-DEC-1993
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 0609.3850003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-087-16

Query Match 16.8%; Score 328; DB 3; Length 595;
Best Local Similarity 27.2%; Pred. No. 1.4e-25;
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

Qy 23 STLERGQKPPPTPSCKLYSIKIQMLDDTQEAPEVQORAPGKVLLDAVCNHLNLVSGDYFG 82
Db 12 SSLKR--KOPKTFTRIVTM-----DAEMEFNCMEKWKGDLDLVGCVRTLGLRETWFFG 63
Qy 83 LEFPDHHKITVWLDLKPIV-KQIRRPKHVVVVKVVFPPD-HTQLOBELTRYLFALQV 140
Db 64 LOY-TIKDTVAVLWKMDKVLHDHVSKEEPVTFHFLAKYPENAEELVQEIQHLLFFLQV 122
Qy 141 KODLAQGRLTCDNTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQ-----D 190
Db 123 KKQILDEKIYCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEBELLKRVINLYQMTPE 182
Qy 191 ALEDKIVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERITAWAEHRGRARDEAEMEYLKIAQDLEMYGVNYFAIRNKKGTTELLIGVDALGLH 242
Qy 251 VFOGFTKIN--AFNNAKVRKLSFKRKRFLIKLRDPDANSAYQDTLEFLMASRDFCKSEWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSKDEFTIK----PLDKIDVFKFNSSKLRVKNKLIQ 298
Qy 308 ICVEHHAFF 316
Db 299 LCIGNHDLF 307
RESULT 12
US-08-179-738-3
; Sequence 3, Application US/08179738
; Patent No. 5578462
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/179,738
; FILING DATE: 10-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
US-08-179-738-3

Query Match 16.8%; Score 328; DB 1; Length 596;
Best Local Similarity 27.2%; Pred. No. 1.4e-25;
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY	23	STLREGOKPPPTSGKLVSTIKIOMQLDQAEFVQAPRGKVLLDVACNHLNLVGDGYFG	82
Db	12	SSLK--KQPKTFTVRITM-----DAEMFNCMKWKGDLDVLVCRTILGIRETWFG	63
QY	83	LEFPDIHKITVWLDDLKPIV-KQIRPRKHVVVKVVFVFPDPD-HTQOQELTRYLFALQV	140
Db	64	LQY-TIKDTVAWLKMDKKVLDHDSKEEPTVTHFLAKFYPENAEELVQOEITQHLLFFLQV	122
QY	141	KODLAQRLTCNDTSAALLISHIVQSEIGDFDEAL--DREHLAKNKVPOQ-----D	190
Db	123	KKOILDEKIVCPPEASVLLASYAVQAQYGDYDPSVHKRGFLAQEELPKRVNLXQMTPE	182
QY	191	ALBCKIVFHHNHIGOTPAESDFOLLEIARRELMYGIRLHPAKDRGRTKINLAVANTGTL	250
Db	183	MWEERTAWTAEBRGARDEAEWEYLKIAQDLEMYGVNFAIRNKGTELLLCVDAALGH	242
QY	251	VFOGFTKIN---AFNMAKVRKLSFKRKREFLIKLRPDANSAYQDTLEFLMASRDFCSFWK	307
Db	243	IYDPENRLTPKISFPNNEIRNISYDKETIK---PLDKKIDVFKFNSSKLRVKNLQI	298
QY	308	ICVEHHAFH	316
Db	299	LCIGNHDLF	307

```

RESULT 13
US-08-628-145-3
; Sequence 3, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kiey, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
US-08-628-145-3

```

```

Query Match          16.8%; Score 328; DB 2; Length 596;
Best Local Similarity 27.2%; Pred. No. 1.4e-25;
Matches      84; Conservative 77; Mismatches 120; Indels 28; Gaps

Qy  23  STLERGQKPPPTSGKLVSIKIOMLDDTQAFEFVQPARAGCKVLLDAYCNHNLNIVEGDYFG 82
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db  12  SSLKR--KQPKTFTVRITVM-----DAEMFECMKWKGDLDLVLCRTLGLRRTWFFG 63
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy  83  LEPPDHKKIIVWLLDLKPIV-KOIRRPKHVVVVFVKFFPPD-HTOLOBELTRYLFALQV 140
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db  64  LQY-TIKDTVAWLKMDKKVLDHVDVKEEPTFFFLAKFFPENAEELVQEIQHLEFFLQV 122
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy  141  KQDLAQGRLLTCDNTSAALLSHIVQSIGDFDBAL-DREHLANKKIYPOO-----D 190
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db  123  KKQILDEKIKYCPPEASVLLASAVAKYQGYDPSVHKRGFLAQEELLPRKRVINLYQMTPE 182
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy  191  ALEDKIIVEFHNNHIGQTPAESDFQLLEIARRLEMYGIRLHPAKDRGRTKINLAVANTGIL 250
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db  183  MWESRITAWYAEHGRARDAEWEYLKIAODLEMYGVNYFAIRNKKGTLELLGVGDALGLH 242
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy  251  VFQGFTRKIN---AFNMAKVKKLSFKRRKRRLKLRPDANSAYQDTLEFLMASRDFCKSFWK 307
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db  243  IYDPENRLTPKISFPWNEIRNISYDKFEFTIK-----PLDKKIDVFKFNSSKLRVKNLILQ 298
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy  308  ICVEHHAFF 316
      |:| |:| |:|
Db  299  LCIGNHDLF 307
      |:| |:| |:|

RESULT 14
US-08-179-738-7
; Sequence 7, Application US/08179738
; Patent No. 5578462
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/179,738
; FILING DATE: 10-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
;
; US-08-179-738-7

```

US-08-179-738-7

```

; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
US-08-628-145-7

Query Match      16.6%; Score 325; DB 1; Length 584;
Best Local Similarity 26.9%; Pred. NO. 2.8e-25;
Matches 83; Conservative 79; Mismatches 119; Indels 28; Gaps 9;

QY 23 STLRCQKPPPTPSGKLVSISKIQMLDDTQAEVFPORAPGKVLDDAVCNHNLNVEGDYFG 82
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKGDLDLVCRTLGLRDTWFFG 63
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 83 LEPPDHKKITVWLDLLKPIV-KQIRRPKHVVVVKVFFPPD-HTQLELTRYLFALQV 140
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 LQY-TIKDTAVLKMCKVLDHVDVSKKEEPTVTHFLAKFYPPENAEELVQEIQTQLFFLQV 122
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 141 KODLAQGRITCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 123 KKEILDEKVCYCPPEASVLLASYAVQAKYGDYDPSVHKRGFLAQEELLPKRVINLYQMTPE 182
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 MWEERITANYAEHRGRARDEAEYELKIAQDLEMYGVNYFTIRNKGTELLGVDALGLH 242
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 251 VFOGFTKIN---AFNNAKVKLSFKRKRFLIKLRPDANSAYQDTLFLMASRDFCKSFWK 307
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 IYDPENRLTPKISFPWNEIRNISYSKFTIK----PLDKKIDVFKFDSSKLRVKNLILQ 298
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 308 ICVEHHAFF 316
   :|:| | | |
Db 299 LCIGNHDLF 307
   :|:| | | |
```

Search completed: December 6, 2001, 08:55:09
Job time: 291 sec

```

Query Match      16.6%; Score 325; DB 1; Length 584;
Best Local Similarity 26.9%; Pred. NO. 2.8e-25;
Matches 83; Conservative 79; Mismatches 119; Indels 28; Gaps 9;

QY 23 STLRCQKPPPTPSGKLVSISKIQMLDDTQAEVFPORAPGKVLDDAVCNHNLNVEGDYFG 82
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKGDLDLVCRTLGLRDTWFFG 63
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 83 LEPPDHKKITVWLDLLKPIV-KQIRRPKHVVVVKVFFPPD-HTQLELTRYLFALQV 140
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 LQY-TIKDTAVLKMCKVLDHVDVSKKEEPTVTHFLAKFYPPENAEELVQEIQTQLFFLQV 122
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 141 KODLAQGRITCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 123 KKEILDEKVCYCPPEASVLLASYAVQAKYGDYDPSVHKRGFLAQEELLPKRVINLYQMTPE 182
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 MWEERITANYAEHRGRARDEAEYELKIAQDLEMYGVNYFTIRNKGTELLGVDALGLH 242
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 251 VFOGFTKIN---AFNNAKVKLSFKRKRFLIKLRPDANSAYQDTLFLMASRDFCKSFWK 307
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 IYDPENRLTPKISFPWNEIRNISYSKFTIK----PLDKKIDVFKFDSSKLRVKNLILQ 298
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 308 ICVEHHAFF 316
   :|:| | | |
Db 299 LCIGNHDLF 307
   :|:| | | |
```

```

RESULT 15
US-08-628-145-7
; Sequence 7, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:52:34 ; Search time 44.17 Seconds
(without alignments)
644.992 Million cell updates/sec

Title: US-09-555-342A-2_COPY_1_374
Perfect score: 1956
Sequence: 1 MGEIQRPTPGSLGAPENS.....GHKKVQFERKHSKIHSIRSL 374

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68:*
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	99.6	1045	2 JC5795	CDEP protein - hum
2	802.5	41.0	1111	2 T23047	hypothetical prote
3	681.5	34.8	850	1 MMH084	erythrocyte membra
4	678.5	34.7	926	1 A41105	protein-tyrosine-p
5	666	34.0	858	2 A46613	protein 4.1, P4.1
6	661.5	33.8	1698	2 T13800	coracine gene prote
7	658	33.6	801	2 A37353	membrane protein 4
8	610	31.2	1026	2 T19631	hypothetical prote
9	527	26.9	568	2 T25859	hypothetical prote
10	506.5	25.9	913	1 A41109	protein-tyrosine-p
11	492.5	23.1	554	2 J00188	membrane protein 4
12	423	21.6	639	2 T20772	hypothetical prote
13	423	21.6	4549	2 T20771	hypothetical prote
14	423	21.6	4667	2 T20771	hypothetical prote
15	368	18.8	1175	2 S51005	protein-tyrosine-p
16	368	18.8	1176	2 T58345	protein tyrosine p
17	362.5	18.5	577	1 A41289	moesin - human
18	361.5	18.5	577	1 S39804	moesin - pig
19	359	18.4	630	2 T47177	hypothetical prote
20	356	18.2	583	1 A46127	radixin - human
21	356	18.2	583	1 S39805	radixin - pig
22	356	18.2	583	1 A41129	radixin - mouse
23	356	18.2	1174	2 T38140	protein-tyrosine-p
24	351	17.9	581	2 T45889	ezrin - bovine
25	349	17.8	586	1 B41129	ezrin - mouse
26	348	17.8	586	1 A34400	ezrin [validated]
27	328	16.8	595	2 S33809	neurofibromin 2 -
28	327.5	16.7	654	2 T30937	hypothetical prote
29	327	16.7	591	2 I54368	merlin protein - m

30	327	16.7	596	2 I68664	merlin - mouse
31	319	16.3	563	2 T29262	hypothetical prote
32	313	16.0	564	2 T29263	hypothetical prote
33	302	15.4	1189	1 JC2366	protein-tyrosine-p
34	300	15.3	1187	1 JC4155	protein-tyrosine-p
35	287	14.7	559	2 A45620	cyto villin homolog
36	286	14.6	559	2 S49143	EG10 protein - tap
37	274	14.0	2490	1 A54971	protein-tyrosine-p
38	268	13.7	2450	2 S71625	protein-tyrosine-p
39	265	13.5	2294	2 I67630	protein tyrosine p
40	265	13.5	2466	2 I67629	protein tyrosine p
41	205	10.5	90	2 A27056	erythrocyte membra
42	155.5	7.9	652	2 T20549	hypothetical prote
43	120	6.1	2491	2 A57036	talain - slime mold
44	116	5.9	2541	2 S11661	talain - mouse
45	106	5.4	444	2 T42674	hypothetical prote

ALIGNMENTS

RESULT 1
JC5795
CDEP protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C:Accession: JC5795
R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.
Biochem. Biophys. Res. Commun. 241, 369-375, 1997
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain
ing factors.
A:Reference number: JC5795; MUID:98086358
A:Accession: JC5795
A:Molecule type: mRNA
A:Residues: 1-1045 <KOY>
A:Cross-references: DBJ:AB008430
C:Comment: This protein is involved in the adhesion, proliferation, and differentiation
C:Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain.homolo
F:1-374/Domain: ezrin-like #status predicted <EZR>
F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>
F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 99.6%; Score 1949; DB 2; Length 1045;
Best Local Similarity 99.7%; Pred. No. 2.le-145;
Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGEIQRPTPGSLGAPENS	GISTLERGOKPPPTPGSGKLVS	IKIOMLDDTQEA	FVPPQRA	60
Db	1	MGEIQRPTPGSLGAPENS	GISTLERGOKPPPTPGSGKLVS	IKIOMLDDTQEA	FVPPQRA	60
Qy	61	PGKVLDDAVCNHNLV	EGDYFGLEFPDHRK	ITVWLDLLKPIVK	QIRRPKHVVVVFYVKFF	120
Db	61	PGKVLDDAVCNHNLV	EGDYFGLEFPDHRK	ITVWLDLLKPIVK	QIRRPKHVVVVFYVKFF	120
Qy	121	PPDHTQLOBELFRYL	FALQVKODLAQGR	LTCNDTSAALISH	IVQSEIGDFEALDREHL	180
Db	121	PPDHTQLOBELFRYL	FALQVKODLAQGR	LTCNDTSAALISH	IVQSEIGDFEALDREHL	180
Qy	181	ANKKVIPODDALEDK	IVFPHHNIHGOTPA	ESDFOLLEIARR	LMYGIIRLHPAKDREGTKI	240
Db	181	ANKKVIPODDALEDK	IVFPHHNIHGOTPA	ESDFOLLEIARR	LMYGIIRLHPAKDREGTKI	240
Qy	241	NLAVANTGILVFGQ	GTKINAFNWKVKLS	FKRKRFLIKLRP	DANSAYQDTLEFLMASRD	300
Db	241	NLAVANTGILVFGQ	GTKINAFNWKVKLS	FKRKRFLIKLRP	DANSAYQDTLEFLMASRD	300
Qy	301	FKCKSEFWKICVEH	AFRLFEPPKPKPV	LFSGSGSFRFSG	RQTQKQVLDYVKEGGHKKVQ	360
Db	301	FKCKSEFWKICVEH	AFRLFEPPKPKPV	LFSGSGSFRFSG	RQTQKQVLDYVKEGGHKKVQ	360
Qy	361	FERKHSKIHSIRSL	374			

Db 361 FERKSHKHSIRSL 374

RESULT 2

T23047

hypothetical protein H05G16.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T23047

R:White, S.

submitted to the EMBL Data Library, June 1997

A:Reference number: T19661

A:Accession: T23047

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-1111 <W1>

A:Cross-references: EMBL:297190; PIDN:CAB10024.1; GSPDB:GN00028; CESP:H05G16.1

A:Experimental source: clone H05G16

C:Genetics:

A:Gene: CESP:H05G16.1

A:Map position: X

A:Introns: 97/3; 139/2; 171/3; 213/2; 294/3; 348/2; 373/2

Query Match

Best Local Similarity 41.0%; Score 802.5; DB 2; Length 1111;

Matches 165; Conservative 64; Mismatches 113; Indels 31; Gaps 6;

QY 22 ISVLERGQKPPPPS-----GKLVSIKIQMLDDTQEAFF----- 55

Db 1 MSNIPRGVAGAPPPGMSNKKRGLVCIKVRMLDDTVAVFHLGVICFFCVSPFHGYVFL 60

QY 56 VPORAPGKVLDAVCNHLNVEGDYGLFEPPOHKKITVWLLDKLPVVKIIR- PKHVVK 114

Db 61 LQKHALGQTLLEDVCRHLNLEDYGLSLFDINGNHCWLDREKTLIRQINGSTDAFY 120

QY 115 FVYFFPPDHTQLQELTRYLFALQVKQDLAOGRLTCNDTSAAALLSHIVQSGIGDF--D 172

Db 121 FVYKFTPNPIDLEEYTRYLFMQIKRDLALGELCHSDNTASLLSAYLVQSECGDFSSE 180

QY 173 EALDREHLAKNKYIQQD-ALEDKIVEFHNNHIGQTPASDQQLLEIARLEMYGRLHP 231

Db 181 DYPDATYLSHTRVPNQTLEFQKKVMDNHRNFTGMTPGESDLAMLEVARRCDFYGVKLA 240

QY 232 AKDREGTKINLAVANTGILVFOGFTKINAFNNAKVKSLFKRKRELKLRDANSAYQDT 291

Db 241 AKDIGNDAALSVMHLGIKVFQRLQDITFSWARIKLSFKRKLKLVKLHPDSYQLKET 300

QY 292 LEFLMASRDFCKSFVKICVEHHAFFRLFEPPKPKPVLF-SRGSFSGRTQKQVLDY 350

Db 301 VEFSPFTRDECKNFKKVEHHAFFRCVQAEPEPKETREFFISKSSFRYHGRQKQLIDY 360

QY 351 VKEGGHKKVQFER 363

Db 361 VREHKKRREFPTR 373

RESULT 3

MMHUE4

erythrocyte membrane protein 4.1, parent splice form - human

N:Contains: erythrocyte membrane protein 4.1; lymphocyte membrane protein 4.1

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1990 #sequence_revision 08-Feb-1996 #text_change 22-Jun-1999

C:Accession: A39810; A60244; A30207; A94143; A31260; B31260; G31260; S13152; A92559; A34

R:Conboy, J.G.; Chan, J.Y.; Chasis, J.A.; Kan, Y.W.; Mohandas, N.

J. Biol. Chem. 266, 8273-8280, 1991

A:Title: Tissue- and development-specific alternative RNA splicing regulates expression

A:Reference number: A39810; MUID:91217063

A:Molecule type: mRNA

A:Residues: 1-850 <CON>

A:Cross-references: GB:IM61733

R:Tang, T.K.; Leto, T.L.; Marchesi, V.T.; Benz Jr., E.J.

Adv. Exp. Med. Biol. 241, 81-95, 1988
A:Title: Expression of specific isoforms of protein 4.1 in erythroid and non-erythroid
A:Reference number: A60244; MUID:89132003
A:Accession: A60244

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-50, 'O', '52', 'L', '54-167', 'F', '169-227', '263-615', '656-787', 'K', '789-850 <TAN>

A:Experimental source: T-cell leukemia line MOLT-4

A>Note: lymphocyte membrane protein 4.1

R:Tang, T.K.; Leto, T.L.; Correia, I.; Alonso, M.A.; Marchesi, V.T.; Benz Jr., E.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3713-3717, 1988

A:Title: Selective expression of an erythroid-specific isoform of protein 4.1.

A:Reference number: A94196; MUID:88234496

A:Accession: A30207

A:Molecule type: mRNA

A:Residues: 210-227, 263-615, 656-787, 'K', '789-850 <TA2>

A:Cross-references: GB:J03796; NID:g182072; PIDN:AAA35794.1; PID:g182074

A:Experimental source: T-cell leukemia line MOLT-4

A>Note: the authors translated the codon AAA for residue 539 as Gln

R:Conboy, J.; Kan, Y.W.; Shohet, S.B.; Mohandas, N.

Proc. Natl. Acad. Sci. U.S.A. 83, 9512-9516, 1986

A:Title: Molecular cloning of protein 4.1, a major structural element of the human er

A:Reference number: A94143; MUID:87092279

A:Accession: A94143

A:Molecule type: mRNA

A:Residues: 210-615, 635-758, 793-850 <CO2>

A:Cross-references: GB:M14993; NID:g182075; PIDN:AAA35795.1; PID:g182076

A:Experimental source: reticulocyte

R:Conboy, J.G.; Chan, J.; Mohandas, N.; Kan, Y.W.

Proc. Natl. Acad. Sci. U.S.A. 85, 9062-9065, 1988

A:Title: Multiple protein 4.1 isoforms produced by alternative splicing in human eryt

A:Reference number: A94215; MUID:89057876

A:Accession: A31260

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 615, 656-700 <CO3>

A:Experimental source: reticulocyte

A:Accession: B31260

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 701-714, 738-837, 'L', '839-848', 'E' <CO4>

A:Experimental source: reticulocyte

A:Accession: C31260

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 701-758, 793-837, 'L', '839-848', 'E' <CO5>

A:Experimental source: reticulocyte

R:Horne, W.C.; Prinz, W.C.; Tang, E.K.Y.

Biochim. Biophys. Acta 1055, 87-92, 1990

A:Title: Identification of two cAMP-dependent phosphorylation sites on erythrocyte pr

A:Reference number: S13152; MUID:91027920

A:Accession: S13152

A:Molecule type: protein

A:Residues: 534-541; 693-701; 758, 793-794 <HOR>

R:Correia, I.; Speicher, D.W.; Marchesi, V.T.

J. Biol. Chem. 261, 13362-13366, 1986

A:Title: Structure of the spectrin-actin binding site of erythrocyte protein 4.1.

A:Reference number: A92559; MUID:87008553

A:Accession: A92559

A:Molecule type: protein

A:Residues: 615, 635-700 <COR>

A>Note: this 67-residue peptide forms a ternary complex with spectrin and actin; anti

A>Note: this peptide contains several possible sites for phosphorylation by CAMP-depe

R:Inaba, M.; Maeda, Y.

J. Biol. Chem. 264, 18149-18155, 1989

A:Title: O-N-acetyl-D-glucosamine moiety on discrete peptide of multiple protein 4.1

A:Reference number: A34377; MUID:90036892

A:Accession: A34377

A:Molecule type: protein

A:Residues: 772-791 <INA>

A>Note: this peptide region appears to have several partially glycosylated Ser and Th

C:Comment: In mammalian erythrocytes, protein 4.1 stabilizes the spectrin-actin netwo

band 3 and glycophorin.

C:Comment: Four domains are proposed based on the properties of peptides released after 3, glycoprotein, and phospholipids. This domain is hydrophobic, and has two potential high in proline.
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase; protein-tyrosine-phosphatase; nonreceptor type 3; GLGF domain homology
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
F:523-597/Domain: GLGF domain homology <GLG>
F:679-900/Domain: protein-tyrosine-phosphatase homology <PTP>
F:852/Active site: Cys (phosphocysteine intermediate) #status predicted
F:858/Binding site: substrate phosphate (Arg) #status predicted

A:Map position: 9q31-q931
A:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase; protein-tyrosine-phosphatase; nonreceptor type 3; GLGF domain homology
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
F:523-597/Domain: GLGF domain homology <GLG>
F:679-900/Domain: protein-tyrosine-phosphatase homology <PTP>
F:852/Active site: Cys (phosphocysteine intermediate) #status predicted
F:858/Binding site: substrate phosphate (Arg) #status predicted

Query Match 34.7%; Score 678.5; DB 1; Length 926;
Best Local Similarity 41.1%; Pred. No. 2.2e-45;
Matches 146; Conservative 62; Mismatches 124; Indels 23; Gaps 6;

QY 35 PSGLKLSIK-----IQMLDDTQEAPEVFPQAPGKVLDDAVCNHNLVEG 78
Db 8 PAGRTYVNRASELARDRQHTVEVCNILLDNTVQAFKVNKHQGVLLDVFVKHLDTQE 67
QY 79 DYFGLFEPDHHKTV-WLDDLKPIVKQIRPKHVHVVFVVKFPPDHTQLOBELTYLFA 137
Db 68 DYFGLQADDSTDNPRWLDPNPKPIKQKRGSPYSLNFRVKFVSDPNKLQEEYTRYQVF 127
QY 138 LQVKDQLAOGRUCLNTSAAALISHIVQSEIGDFDEALDR-HLAKNKYIPQO-DALEK 195
Db 128 LQVKDILGRPCPSNTAALLASFAVQSELGDDYDQSENLSGLSYDYSFIPNQPDPEKE 187
QY 196 IVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGF 255
Db 188 IAKHQHIGLSPAEAEFNLARTLELYGVFEHYFVARDQSNNEIMIGVMSGGILYKVR 247
QY 256 TKINAFNAKVKRLSPKRRFLIKLRPDANSAYODTLEFLMASRDFCKFWKICVBEHAF 315
Db 248 VRMTFPPKLVKISPKCKOFFQLKELHESRETLGFMVNYRACKNLWACVBEHHTF 307
QY 316 FRLFEPEPKPKPVL---FSRGSFSPFSGRTQKQVLDYVKEGGHKKVQPERKHSK 367
Db 308 FRL-DRPLPPQKNFFAHYFTLGSKFYCGRTVEQVQYKGEKANKDRVFARSFK 361

RESULT 5
A46613
protein 4.1, P4.1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 25-Aug-1995
C:Accession: A46613
J: Huang, J.P.; Tang, C.J.; Kou, G.H.; Marchesi, V.T.; Benz Jr., E.J.; Tang, T.K.
J: Biol. Chem. 268, 3758-3766, 1993
A:Title: Genomic structure of the locus encoding protein 4.1. Structural basis for co
A:Reference number: A46613; MUID:93155238
A:Accession: A46613
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-858 <HUA>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBI:124466, NCBI:124467)
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
F:213-488/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 34.0%; Score 666; DB 2; Length 858;
Best Local Similarity 40.9%; Pred. No. 1.9e-44;
Matches 158; Conservative 53; Mismatches 155; Indels 20; Gaps 9;

QY 5 EORPTPGSRL---GAPENSGISTLE---RGOKPP-----PTPSGKLVIKIQMLDDTQ 51
Db 163 EHREDPDSFETKEGEGIECSGTEVKEDPSRAEREPEASQKPVRRHNRHCKVSLDDTV 222
QY 52 EAFEPQAPGKVLDDAVCNHNLVEGDFGLFEPDHHKTVWLDLKLPIVKQIRR-PKH 110
Db 223 YECVVEKHANLQDLKRVCEHLNLEEDYFGLALWDSATSKTLDLSAKETKKQVGPWN 282
QY 111 VVVKFVVFPPDHTQLOBELTRYLFAVQKQDLAOLGRLTCNDTSALLISHIVQSEIGD 170

Query Match 34.8%; Score 681.5; DB 1; Length 850;
Best Local Similarity 42.0%; Pred. No. 1.1e-45;
Matches 152; Conservative 58; Mismatches 143; Indels 9; Gaps 7;

QY 16 APENSGISTLERGQPPPTPSGLKLSIKIQMLDDTQEAPEVFPQAPGKVLDDAVCNHNL 75
Db 188 SPQSKAETELKASQK--PIRKHRNHCKVSLDDTVYECVWEKHAQGVLLKRVCEHLNL 245
QY 76 VEGDYFGLFEPDHHKTVWLDLKLPIVKQIRR-PKHVVVVFVVKFPPDHTQLOBELTRY 134
Db 246 LLEDYFGLAIWNAATSKTWLSAKETKKQVGPWNFT--FNVFPPPPAQLTEDITRY 303
QY 135 LFAVQKQDLAOLGRLTCNDTSAAALISHIVQSEIGDFDEALDR-EHLAKNKYIPQO-DAL 192
Db 304 YLCQLRQDIVAGRPCSFATLALLGSYTIQSELGDDYDPELHGVDYVDFKLAPNOTKEL 363
QY 193 EDKIVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILV 252
Db 364 EKVVELHKSYSRMTPAQADLEFLENNAKLSMYGVDLHKAKDLEGVDITLGVCSGLLYV 423
QY 253 OGFTKINAFNAKVKRLSPKRRFLIKLRPDANSAYODTLEFLMASRDFCKFWKICVBEH 312
Db 424 KDKLINRPPKVLKISYKSFIFIKIRPGEQYETISGFKLPSYRAAKLWVCEH 483
QY 313 HAFRLFEPEPKPKPVLSRGSFSPFSGRTQKQVLDYVKEGGHKKVQPERKHSKIHSTR 372
Db 484 HFFRL-TSTDIPKSKFLALGSKFRYSRTQAOTQASALIDRPAHPERTASK-RASR 541
QY 373 SL 374
Db 542 SL 543

RESULT 4
A41105
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 4 - human
N:Alternate names: PTPase MEG
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A41105
R:Gu, M.; York, J.D.; Warshawsky, I.; Majerus, P.W.
Proc. Natl. Acad. Sci. U.S.A. 88, 5867-5871, 1991
A:Title: Identification, cloning, and expression of a cytosolic megakaryocyte protein-ty
A:Reference number: A41105; MUID:91288564
A:Accession: A41105
A:Molecule type: mRNA
A:Residues: 1-926 <GUA>
A:Cross-references: GB:M68941; NID:g190747; PIDN:AAA36530.1; PID:g190748
A:Experimental source: megakaryocytes, cell line MEG-10
C:Genetics:
A:Gene: GDB:PTPM4
A:Cross-references: GDB:131387; OMIM:176878

Db 283 FT--FNVKFPDPDAQITEDITRYIYLCQLQRODIVAGRLPCSFATLALLGSYTTIQSELGD 340

Qy 171 FDEAL-DREHLAKNKYIPQO-DALEDKIVFEFHNNHIGOTPAESDFOLLEIARRLMEYGR 228

Db 341 YDPELHGMDYVDSFKLAPNQTRELEEKVMELHKSYSRMTPAQADLEFLNNAKKLSMYGVD 400

Qy 229 LHPAKDREGTKINLAVANTGILVFOGFTKINAFNNAWKVLSFKRKREFLIKLRPDANSAY 288

Db 401 LHKAKDLEGVDIILGVCSSGLLYWKLKLRINRFPWPKVLTKSSKRSSFYKIRPGEQEHY 460

Qy 289 QDTLEFLMASRDFCKSFWKICVHEHFAFFLEEPKPKPVLFSRGSFSGRTQKQVL 348

Db 461 ESTIGFKLPSYRAAKLKWVCVEHHFFRL-TSTDTIPKSFALUGSKFRYSGRTOAQTR 519

Qy 349 DYVKEGHHKKVQPERKSHKHSIRSL 374

Db 520 QASALIDRPAPHFERTASK-RASRSL 544

RESULT 6

T13800

coracle gene protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13800

R:Feihon, R.G.; Dawson, I.A.; Artavanis-Tsakonas, S.

A:Title: A *Drosophila* homologue of membrane-skeleton protein 4.1 is associated with the development of the nervous system

A:Reference number: Z17763; MUID:94215495

A:Accession: T13800

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1698 <FEH>

A:Cross-references: EMBL:L27467; NID:g440293; PID:g440294; PIDN:AAB59187.1

C:Genetics:

A:Gene: coracle

A:Cross-references: FlyBase:FBgn0010434

A:Map position: 2

C:Function:

A:Description: may play a role in cell-cell interactions

C:Keywords: alternative splicing

Query Match 33.8%; Score 661.5; DB 2; Length 1698;

Best Local Similarity 40.8%; Pred. No. 1.1e-43;

Matches 142; Conservative 64; Mismatches 125; Indels 17; Gaps

Qy 5 EQRPFGSRUGAPENGSGISTLKGOKPPPTPS--GKLVSIKIQLMDDTQAEFVQPARGP 62

Db 4 EIKPSAPAEPTPTKS-----KPKSSSSHGKPALARVTLLDGLSDVSDRKAIG 54

Qy 63 KVLLDVAVCNHLNVEGDYFGLGEPDHHKITVWLDLLKPLVKQIRPKHVVKVVFPP 122

Db 55 RDVINSIAGLNIEKDYFGLTYETPTDPRWLDLEKPVSKFFRDVTWPLT-FAVKYTPP 113

Qy 123 DHTQLOELTRYLFALQVQKDLAQGRLTCDNTSAALLSHIVQSGIDFD--EALDREHL 180

Db 114 EPSQLKEDITRYHLCLQVRNDILEGLPCTFFYTHALLGSYLVSQEMGYDDEEPTRAYL 173

Qy 181 AKNKYIPQODA-LEDKIVFEFHNNHIGOTPAESDFOLLEIARRLMEYGRILHPAKDREGTK 239

Db 174 KDEFIAPNQTALEDKVMDLHKTGQSPAELHYLENNAKKLAMYGVLDLHPAKDSEGV 233

Qy 240 INLAVANTGILVFOGFTKINAFNNAWKVLSFKRKREFLIKLRPDANSAYQDTLEFLMASR 299

Db 234 IMLGVCSSGLLYWKLKLRINRFPWPKVLTKSSKRSSFYKIRPGEQEHY 293

Qy 300 DFCKSFWKICVHEHFAFFLEEPKPKPVLFS-RGSSSFSGRTQKQ 346

Db 294 RAKKLWKSVEHHTFRLM-TPEVPVSKMFPVFGSYIRYKGRTOAE 340

RESULT 7

A37353
membrane protein 4.1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 13-Aug-1999
C:Accession: A37353; A29901
R:Spencer, M.; Giebelhaus, D.H.; Kelly, G.M.; Bicknell, J.; Florio, S.K.; Milam, A.H.
Dev. Biol. 139, 279-291, 1990
A>Title: Membrane skeleton protein 4.1 in developing Xenopus: expression in postmitot
A:Reference number: A37353; MUID:90249600
A:Accession: A37353
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-801 <SPE>
A:Cross-references: GB:M20621; NID:g214090; PIDN:AAA49695.1; PID:g214091
R:Giebelhaus, D.H.; Eib, D.W.; Moon, R.T.
Cell 53, 601-615, 1988
A>Title: Antisense RNA inhibits expression of membrane skeleton protein 4.1 during em
A:Reference number: A29901; MUID:88223353
A:Accession: A29901
A:Molecule type: mRNA
A:Residues: 1-550 <GIE>
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
C:Keywords: cytoskeleton; membrane protein
F:195-470/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 33.5%; Score 658; DB 2; Length 801;
Best Local Similarity 41.5%; Pred.No. 7.4e-44;
Matches 139; Conservative 60; Mismatches 116; Indels 20; Gaps 6;

Qy 28 GKPKPPPTSGKL-----VSIKIMLDTPQAFEPVRAPGVLLDVCNNHNL 75
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 169 GKEPIKKPEGSKASHKVRRSPNMRCVKTLDDTVYECLEKHAKGDIFKKVCSHLNI 228

Qy 76 VEGDYFGLEFPDHKKITVLDLLKPIVQIRRKHVVF--VKFPDPHTOLQELTR 133
|| |||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 229 VEEDYFLGAIWESPTCRVWLDPDLKRQVHGPG---CEFTSNVKFYPPDPAQLSEDIR 285

Qy 134 YLFALQVKODLAQCRLCNDSATALLISHIVQSIGDFDEALDR-EHLAKNKVIPOODA- 191
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 286 YYLCQLRKDI FSRGLPCSFATLALGSYTVQSEVGDEEELHGVVDYVSEFKLSPNQTKD 345

Qy 192 LEDKIVEFHNNHIQTPAESDFQLLEIARMEVGIIRLHPAKDREGTKINLANVTGILV 251
||::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 346 LEEKVGELHKYSRWMTQAQADLEFLNAKKLTMYGVDIHQAOKLEGVDIKLGVCSSGLMV 405

Qy 252 FQGTFKNAPFWAVKRLSPKRKFLLKLPDANSAYQDTLEFLMASRDCKSFWKICVE 311
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 406 FKDLNRINRPWPVKISYKRSFFIKIRPGEQEYSTIGFKLPSYKAACKLWKVCVE 465

Qy 312 HHAFFRLFEPEKPKPVLSRGSSFRFSGRTKOQ 346
|| |||| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 466 HHTFRL-TSTESIPKHFRLSLGSTFRYSRTQAQ 499

RESULT 8
T19631
hypothetical protein C48D5.2a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T19631; T20040
R:Thomas, K.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19154
A:Accession: T19631
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1026 <WIL>
A:Cross-references: EMBL:Z48241; PIDN:CXA88287.1; GSPDB:GN00021; CESP:C48D5.2a
R:Lightning, J.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19215

A:Accession: T20040
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1026 <W12>
A:Cross-references: EMBL:U36237; PIDN:CAAB5272.1; GSPDB:GN00021; CESP:C48D5.2a
A:Experimental source: clone C48D5
C:Genetics:
A:Gene: CESP:C48D5.2a
A:Map position: 3
A:Introns: 46/3; 158/1; 200/2; 282/3; 358/2; 396/2; 447/2; 481/3; 543/1; 605/3; 70/2
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; P

Query Match 31.2%; Score 610; DB 2; Length 1026;
Best Local Similarity 39.1%; Pred. No. 6.3e-40;
Matches 147; Conservative 60; Mismatches 145; Indels 24; Gaps 10;

QY 13 RLGAPEISGISTLERQKPPPTPSGKLVSIKIQLMDLDTQAEFVQAPACKVLLDAVCNH 72
Db 2 RLGSNSYDQRTAIGQTPVTPPPNQIRCTVTFDLSTSYHFEIEKNSLGIVLLEKVFNY 61
QY 73 LNLVEGDYFGLGF-----PDHKKITVWLDLLKPIVKO-IRPKHVVVVKFVVKFPPDPH 124
Db 62 LEIEKDYEGLVFIADVNSAQKK---NLDPSKLNLRKQMICPPYHLF--FRVKFVVRDP 116
QY 125 TQLOEELRYLFAIQKDLAQRITCNDTSAALLISHIVQSEIGDFDE---ALDREHLA 181
Db 117 NLRDEFTRFQYQVQRNLEGRLEPCNGLSALLASYVYVQAEVGFEEKTHGMSRTCLC 176
QY 182 -KNKIYPOQDALEKIVFHHNHIGQTPAESDFOLLEIARRHLEMYGIRLHPAKDREGTKI 240
Db 177 YKIQFATLPDSDRVAELHQLHIGQTPDVAEQNFLDHARRLEMYGMDVYGDVANDHLPI 236
QY 241 NLAVANTGILVFGFTKINAFNMAKVRKLSFKRKLRLKLRPDANSAYQDTLEFLMASRD 300
Db 237 EIGVGAVGKIFHEGIMKMEYAVIRKLSFKKKQVQ-VLANEDGVSETIMIFNMSAK 295
QY 301 FCKSFWKICVHHAFRLPEEPKPKPVLFSGSSFRSGRTQKQVL---DYVKEGGHK 357
Db 296 ICKLLKMKCIEQHTFRLKTPPK-TPQRKVFNFSGFRYSGRTEYQTLLENEHRKSAGHR 354
QY 358 KVOFERKSHKSIHS 373
Db 355 --NFRSLSKSFLRS 368

RESULT 9
T25859
hypothetical protein T04C9.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25859
R:Favella, A.
submitted to The EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid T04C9.
A:Reference number: Z20101
A:Accession: T25859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-568 <FAV>
A:Cross-references: EMBL:U80955; PIDN:AAB38104.1; GSPDB:GN00021; CESP:T04C9.6
A:Experimental source: strain Bristol N2; clone T04C9
C:Genetics:
A:Gene: CESP:T04C9.6
A:Map position: 3
A:Introns: 52/3; 87/3; 129/2; 162/1; 197/2; 280/3; 393/3; 538/3

Query Match 26.9%; Score 527; DB 2; Length 568;
Best Local Similarity 35.9%; Pred. No. 9.7e-34;
Matches 125; Conservative 64; Mismatches 139; Indels 20; Gaps 8;

QY 38 KLVSIKIQMLDQTQAEFVQAPACKVLLDAVCNHLNLVEGDYFGLFDPDHKKITVWLDL 97

Db 33 KFOCKVLLLDGAHLNLVVPNRAVGSELVEEYFYSLDLEEDYFGLQYTDFFNVQHWLDP 92
QY 98 LKPIVKQIRRPKHVVVVKVVFPPD-HTQLQELRYLFAIQKDLAQRITCNDTSA 156
Db 93 TKKVAQVAIGPPFTLRFVRKFTSPSSNLKLELRYQFFLQIKQDISSGRQCQPHLA 152
QY 157 ALLISHIVQSEIGDFDEALDRE-HLAKNKYIPOQD-ALEDKIVFHHNHIGQTPAESDFQ 214
Db 153 IELAAFLASGLDYNPELHTALFISEFRPHPEQDEKMEVEILE-----SGQTPAQALN 207
QY 215 LLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNMAKVRKLSFKR 274
Db 208 YLNKARWIEYGVDMHIVEGKDGNTYRLGLTPQGLVLFDPQKIGLFEKLEKQLDFKKN 267
QY 275 RELIKLRPDANSAYQ-----DTLEFLMASRDFCKSEKFWKICVHHAFRLPEEPKPKPV 329
Db 268 KITLVEEDADQSNNGIQIQLHTFVHLTSEKAAKHEWKCAIEQHAFFRLKSRPQIANRKI 327
QY 330 LFSR-CSSFRFSRGTOKQVLDYVKEG---HKKVQFERKSHKSIHSIR 372
Db 328 QFFRLGSTFKYRGRTYEYTIH--KEGARLSRRQSCSFERRPSQRYGPR 373

RESULT 10
A41109
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human
N:Alternate names: PTPH1
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A41109; I55698
R:Yang, Q.; Tonks, N.K.
Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991
A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with
A:Reference number: A41109; MUID:91296738
A:Accession: A41109
A:Molecule type: mRNA
A:Residues: 1-913 <FAN>
A:Cross-references: GB:I64572; NID:9179912; PIDN:AAA35647.1; PID:9179913
R:Ikuta, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.
J. Gastroenterol. 29, 727-732, 1994
A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA
A:Reference number: I55698; MUID:95179278
A:Accession: I55698
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 899-913 <RES>
A:Cross-references: GB:S76309; NID:913165; PIDN:AAB33583.1; PID:913166
C:Genetics:
A:Gene: GDB:PTPN3
A:Cross-references: GDB:I31386; OMIM:176877
A:Map position: 9q31-9q31
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
F:516-590/Domain: GLGF domain homology <GLG>
F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
F:842/Active site: Cys (phosphocysteine intermediate) #status predicted
F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 25.9%; Score 506.5; DB 1; Length 913;
Best Local Similarity 35.0%; Pred. No. 7.6e-32;
Matches 119; Conservative 68; Mismatches 134; Indels 19; Gaps 9;

QY 27 RGOKPPPTPSGKLVSIKIQLMDLDTQAEFVQAPACKVLLDAVCNHLNLVEGDYFGLFEP 86
Db 16 RYSEKPKRTRSEVICSIHFLGVGVTFKVTQDTCQVLDVMVHNLGVTEKEFYFGLQHD 75
QY 87 DHKKITV-WLDLLKPIVKQIRPKHVVVVKFVVFPPDPHTQLOEELRYLFAIQKDLA 145
Db 76 DSDVDSFRLWEASKPIRKQLKGGFPCTLHFRVRFPPDPNTLQEQTRHLYFLQLKWDIC 135

QY 146 QGRLTNDTSAALLSHIVQSEIGDDEALDRE-HLAKNKYIPQDALED---KIVEFHH 201
Db 136 EGRLTCPNLSAVVASYVQSHFGYNSHTHPGYLSDSHFIPDQN--EDFLTKVESLHE 193
QY 202 NHIGQTPAESDFQLLETARLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTKINAF 261
Db 194 QHSGLKQSEAEYCINIARTLDYGVYELHSGRDLHNLDMIGTASGAVAYRYKICTSY 253
QY 262 NNAKVRKLSKRRKRIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFRLLEE 321
Db 254 PWNILKISFKRKKEFIHQKQKQAESREHIVAFNMLNRYCKNLKSCVEHHTFFQA-KK 312
QY 322 PKPKPKPVL---PSRGSFRSFGRTQKQVLD-VYKE--GG 355
Db 313 LLPQKNVLSQYWTMGSR-----NTRKSVNNQCKKVIIG 347
RESULT 11
J00188
membrane protein 4.1 homolog - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Feb-1994 #sequence_revision 20-Aug-1994 #text_change 05-Nov-1999
C:Accession: I55505; J00188
R:Takeuchi, K.; Kawashima, A.; Nagafuchi, A.; Tsukita, S.
J. Cell Sci. 107, 1921-1928, 1994
A:Title: Structural diversity of band4.1 superfamily members.
A:Reference number: I55505; MUID:95074267
A:Accession: I55505
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-534 <RES>
A:Cross-references: GB:D28818; NID:9466547; PIDN:BAA05978.1; PID:9466548
C:Superfamily: protein 4.1 membrane-binding domain homology
F:13-295/Domain: protein 4.1 membrane-binding domain homology <B41>
Query Match 23.1%; Score 452.5; DB 2; Length 554;
Best Local Similarity 33.1%; Pred. No. 6.9e-28;
Matches 105; Conservative 65; Mismatches 132; Indels 15; Gaps 7;
QY 43 KIOMLDDTQAEFVPQR-----APGKVLDDAVCNHNLNVEGDYFGLFDPDHKKITVWLDL 97
Db 14 EVLLDESLLTLTQOQGIKSTKGSVLDHVRHNLNVEIDYFGLRYCDRSHQTYWLDLP 73
QY 98 LKPIVQK---IRPKHVVVKKVFPDPDHTLOELTRYLAFALQVKDLAGOGLTCNDT 154
Db 74 AKTLAEKELINTGPPYTLVFGIKFAEDPCKLKEITRYQFLQVKDALOGLRCPVYN 133
QY 155 SAALLSHIVQSIGDFDEALDRE-HLAKNKYIP-QQDALEDKIVFHHNHIGQTPAESD 212
Db 134 IAAQMGAYAIQABLDHPYKHTAGVYSEVRFVDPQKEELEAIERIHKTLMGQAPSEAE 193
QY 213 FOLLETARLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTKINAFNNAKVRKLSFK 272
Db 194 LNLRYTAKSLEMYGVDLHPYGENKSEYFGLTPSGVVYVKKQKGYFWPRITKVFHK 253
QY 273 RKRLIK-LRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFRLFEKPKPKPVLF 331
Db 254 ETQFELRVLGKDCN---EFSFFEASKTACKHLWKSCVEHHTFFRMPDTESNLSRKLS 310
QY 332 SRGS-SFRSFGRTQKV 347
Db 311 KFGSISYKHYRTALQM 327
RESULT 12
J20772
hypothetical protein ZK270.2b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20772; T27815
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19322
A:Accession: T20772
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-639 <WIL>
A:Cross-references: EMBL:Z81499; PIDN:CAB04091.1; GSPDB:GN00019; CESP:ZK270.2b
A:Experimental source: clone FLIC3
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20423
A:Accession: T27815
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-639 <W12>
A:Cross-references: EMBL:Z82089; PIDN:CAB05004.1; GSPDB:GN00019; CESP:ZK270.2b
A:Experimental source: clone ZK270
C:Genetics:
A:Gene: CESP:ZK270.2b
A:Map position: 1
A:Introns: 61/3; 112/1; 244/3; 328/2; 384/1; 433/3; 467/2; 559/3; 610/3
Query Match 21.6%; Score 423; DB 2; Length 639;
Best Local Similarity 33.8%; Pred. No. 1.8e-25;
Matches 113; Conservative 58; Mismatches 145; Indels 18; Gaps 10;
QY 28 GOKPPPTPS-----GRLVSIKIQMLDDTQAEFVPQAPGKVLDDAVCNHNLNVEGDYFG 82
Db 27 GQKHPTTSQQRDSKMLAKVLLPDGVQKREFEVNKNSEGEALFROVTRDLSIEEYFS 86
QY 83 LEFPDHHKIT-VMLDILLKPIVKQIR-RPKHVVVKKVFPDPDHTLOELTRYLAFALQV 140
Db 87 LCFYDKDEGTRHLYNDKNIAQIKGLPWE--FSPEVKFYPTPTTIVDDHARYVYVFLQL 144
QY 141 KQDLAAGRLTCNDTSAALLSHIVQSEIGDFD-EALD--REHLAKNKYIPQDALED--- 194
Db 145 RDLTLTGRLPATADTSHLSGSFVAQIEFGDAPAEMTDAYEQFIVASKLPVSAQAPETVK 204
QY 195 KIVEFHNNHIGQTPAESDFOLLEIARRLEMYGIRLHPA-KDREGTKINLAVANTGILVQ 253
Db 205 KIVDLHREMRGQTPSEAEQDFLDHCKHLALYGIHLFKALSDKDKKVPDVGVIGAGINIQ 264
QY 254 GFTKINAFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHH 313
Db 265 DEQKTHSFQWNLIKIGYRRTVFSIKKAGTVEKNEKTLVFKLPNHVAAKRTWKCAVEHH 324
QY 314 AFRLFEKPKPKPKPVLSRGS--FRFSGRTOQK 346
Db 325 TFFRLT-QPEDKTHKSFNFQSGRFRYQGRTOFQ 357
RESULT 13
J20771
hypothetical protein ZK270.2a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20771; T27814
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19322
A:Accession: T20771
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-4549 <WIL>
A:Cross-references: EMBL:Z81499; PIDN:CAB04090.1; GSPDB:GN00019; CESP:ZK270.2a
A:Experimental source: clone FLIC3
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20423
A:Accession: T27814
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-4549 <W12>

Db 186 QKVALHQQYRGTLTAPAEMLYMOEVERMDGYGEESYPKADSQGSDISIGACLDGIFVKH 245
Qy 252 FQGETKINAFNWKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSEFWKICVE 311
Db 246 KNGRPPV-VFRWHDIANNSHNSFFALEL--ANK--EETIQOTEDMETAKYVWRLCVA 299
Qy 312 HHAFRLFEFPKPKPKPVLS--RGSSFRFS 340
Db 300 RHKEYRLNQCNLQTOAATLNSVRRGSSSRMS 330

Search completed: December 6, 2001, 08:52:37
Job time: 369 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:53:07 ; Search time 28.69 Seconds
(without alignments)
477.959 Million cell updates/sec

Title: US-09-555-342A-2_COPY_1_374

Perfect score: 1956

Sequence: 1 MGEIEQRPTGSRIGAPENS.....GHKKVQFERKXKIHISRLS 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681.5	34.8	864	1 41_HUMAN	P11171 homo sapien
2	678.5	34.7	926	1 PTV4_HUMAN	P29074 homo sapien
3	669.5	34.2	881	1 E4L1_HUMAN	Q9H490 homo sapien
4	666	34.0	858	1 41_MOUSE	P48193 mus musculus
5	658	33.6	801	1 41_XENLA	P11434 xenopus lae
6	610	31.2	1026	1 PTL1_CAEEL	P28191 caenorhabdi
7	587.5	30.0	732	1 YF48_HUMAN	Q9H0M4 homo sapien
8	506.5	25.9	913	1 PTN3_HUMAN	P26045 homo sapien
9	496	25.4	619	1 NBL4_BRARE	O57457 brachydanio
10	480.5	24.6	598	1 NBL4_HUMAN	Q9HCS5 homo sapien
11	452.5	23.1	554	1 NBL4_MOUSE	P52963 mus musculus
12	368	18.8	1175	1 PTNL_RAT	Q62728 rattus norv
13	368	18.8	1176	1 PTNL_MOUSE	Q62136 mus musculus
14	363.5	18.6	576	1 MOES_MOUSE	P26041 mus musculus
15	362.5	18.5	576	1 MOES_HUMAN	P26038 homo sapien
16	361.5	18.5	576	1 MOES_PIG	P26042 sus scrofa
17	356	18.2	583	1 RADI_HUMAN	P35241 homo sapien
18	356	18.2	583	1 RADI_MOUSE	P26043 mus musculus
19	356	18.2	583	1 RADI_PIG	P26044 sus scrofa
20	356	18.2	1174	1 PTNL_HUMAN	Q16825 homo sapien
21	351	17.9	580	1 EZRI_BOVIN	P31976 bos taurus
22	349	17.8	585	1 EZRI_MOUSE	P26040 mus musculus
23	348	17.8	585	1 EZRI_HUMAN	P15311 homo sapien
24	328	16.8	595	1 MERL_HUMAN	P35240 homo sapien
25	327	16.7	596	1 MERL_MOUSE	P46662 mus musculus
26	306.5	15.7	572	1 MOES_LYTA	P52962 lytechinus
27	302	15.4	1189	1 PTNE_MOUSE	Q62130 mus musculus
28	300	15.3	578	1 MOEH_DROME	P46150 drosophila
29	300	15.3	1187	1 PTNE_HUMAN	Q15678 homo sapien
30	276.5	14.1	1051	1 YC94_HUMAN	Q9P292 homo sapien
31	265	13.5	2485	1 PTND_HUMAN	Q12523 homo sapien
32	251	12.8	1062	1 YAI3_HUMAN	Q9Y216 homo sapien
33	205	10.5	90	1 YAI1_CHICK	P12264 gallus gall

RESULT 1

41_HUMAN	6.1	2492	1	TALA_DICDI	P54633 dictyosteli
34	120	6.1	2492	1 TALI_MOUSE	Q9Y490 homo sapien
35	119	6.1	2541	1 TALI_MOUSE	P26039 mus musculus
36	116	5.9	2541	1 TALI_MOUSE	Q07436 drosophila
37	106	5.4	1429	1 EXPA_DROME	Q13402 homo sapien
38	98.5	5.0	2215	1 MY7A_HUMAN	P97479 mus musculus
39	96.5	4.9	2215	1 MY7A_MOUSE	P87185 candida alb
40	94	4.8	488	1 NFSI_CANAL	Q13596 homo sapien
41	93	4.8	522	1 SNX1_HUMAN	Q61001 mus musculus
42	92.5	4.7	3635	1 LMA5_MOUSE	O55439 bovine aden
43	91.5	4.7	663	1 TERM_ADEB3	P87187 candida mal
44	91	4.7	484	1 NFSI_CANMA	Q17391 caenorhabdi
45	91	4.7	780	1 CUL3_CAEEL	

ALIGNMENTS

41_HUMAN	STANDARD;	PRT;	864 AA.
AC	P11171; P11176; Q9Y578; Q9Y579; Q14245;		
DT	01-JUL-1989 (Rel. 11, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	PROTEIN 4.1 (BAND 4.1) (P4.1) (EPB4.1).		
GN	EPB41 OR E41P.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).		
RX	MEDLINE=88234496; PubMed=3375238;		
RA	Tang T.K., Leto T.L., Correas I., Alonso M.A., Marchesi V.T.,		
RA	Benz E.J. Jr.;		
RT	"Selective expression of an erythroid-specific isoform of protein		
RT	4.1.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 85:3713-3717(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).		
RX	MEDLINE=89132003; PubMed=3223413;		
RA	Tang T.K., Leto T.L., Marchesi V.T., Benz E.J. Jr.;		
RT	"Expression of specific isoforms of protein 4.1 in erythroid and non-		
RT	erythroid tissues.";		
RL	Adv. Exp. Med. Biol. 241:81-95(1988).		
RN	[3]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 378-393 (ERYTHROID ISOFORM).		
RC	TISSUE=Reticulocytes;		
RX	MEDLINE=87092279; PubMed=3467321;		
RA	Conboy J., Kan Y.W., Shohet S.B., Mohandas N.;		
RT	"Molecular cloning of protein 4.1, a major structural element of the		
RT	human erythrocyte membrane skeleton.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 83:9512-9516(1986).		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM 3).		
RX	MEDLINE=91217063; PubMed=2022644;		
RA	Conboy J.G., Chan J.Y.C., Chasis J.A., Kan Y.W., Mohandas N.;		
RT	"Tissue- and development-specific alternative RNA splicing regulates		
RT	expression of multiple isoforms of erythroid membrane protein 4.1.";		
RL	J. Biol. Chem. 266:8273-8280(1991).		
RN	[5]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RX	Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;		
RA	"Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5.";		
RT	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[6]		
RP	SEQUENCE OF 157-227 FROM N.A., AND VARIANT ILE-214.		
RX	Lichtenauer U., Huang S.C., Vortmeyer A., Zhuang Z.;		
RT	"Valine to isoleucine polymorphism in exon 4 of human protein 4.1.";		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[7]		
RP	SEQUENCE OF 648-714.		
RX	MEDLINE=87008553; PubMed=3531202;		

RT "Identification, cloning, and expression of a cytosolic megakaryocyte
 RT protein-tyrosine-phosphatase with sequence homology to cytoskeletal
 RL protein 4.1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5867-5871(1991).
 CC -!- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
 CC CYTOSKELETON.
 CC
 CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATASE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC
 CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC
 CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC
 CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M68941; AAA36530.1; -;
 CC PIR; A41105; A41105.
 CC HSP; P18052; LYFO.
 CC MIM; 176878; -;
 CC InterPro; IPR000299; Band_4.1.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR000387; TYR_phosphatase.
 CC InterPro; IPR000242; TYR_prot_phptase.
 CC Pfam; PF00373; Band_41; 1.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF00102; Y_phosphatase; 1.
 CC PRINTS; PR00700; PRTYPHPTASE.
 CC PRINTS; PR00935; BAND41.
 CC SMART; SM00295; B41; 1.
 CC SMART; SM00228; PDZ; 1.
 CC SMART; SM00194; FPPC; 1.
 CC PROSITE; PS00660; BAND_41_1; 1.
 CC PROSITE; PS00661; BAND_41_2; 1.
 CC PROSITE; PS50057; BAND_41_3; 1.
 CC PROSITE; PS50106; PDZ; 1.
 CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 CC PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 CC PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 CC Structural protein; Cytoskeleton; Hydrolase.
 KW DOMAIN 84 241 BAND 4.1-LIKE.
 FT
 FT DOMAIN 517 589 PDZ.
 FT DOMAIN 677 926 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 852 BY SIMILARITY.
 SQ SEQUENCE 926 AA; 105911 MW; 4DAC6A87A675CFB0 CRC64;
 Query Match 34.7%; Score 678.5; DB 1; Length 926;
 Best Local Similarity 41.1%; Pred. No. 1.7e-46;
 Matches 146; Conservative 62; Mismatches 124; Indels 23; Gaps 6;
 QY 35 PSGLKVSIR -IQMLDDTQEAPEVQAPGKVLDDAVCNHNLVLEG 78
 Db 8 PAGRYNVVASELARDROHTEVCNILLDDNTVQAEKVNKHQGVLLDVVFKHLDTQ 67
 QY 79 DYFGLFPPDHKITV-WDLLKPIVKQIRPKHVVFVVKVFPFPPDHTQLQELTYLFA 137
 Db 68 DYFGLQLADDSDNPRWLDNPNKPIRQLKRGSPYSLNFRVKFVSPNKLQEYTRYQYF 127
 QY 138 LQVKODLAOGLTCDNTSAAALISHTVQSEIGDFDEALDRE-HLAKNKYIPQO-DALEDK 195
 Db 128 LQIKODILTRLPSCNNTAALLASFAVQSELGSDYDQSENLGSLDSYSPINQPODFEKE 187
 QY 196 IVEFHNHIGQTPAESDFOLLETARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGF 255
 Db 188 IAKLHQHGLGLSPAEEFNLTARTLEYGVFFHVARQDSNNEIMVWSGGILYKNR 247
 QY 256 TTINAFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHFAF 315

Db 248 VRNTPFWLKIIVKISFKCKQFFIQLRKELHESRETLLGFNMVNYRACKNLWKACVHHFF 307
 QY 316 FRLFEPPKPKPVL---FSRGSSFRFSQRTQKQVLDYVKEGHHKKVQFERKHSK 367
 Db 308 FRL-DRPLPPQKNFFAHYFTLGSKFRYCGRTEVSQVQYKGRKANKDRVFARSPSK 361
 RESULT 3
 E4LL_HUMAN
 ID E4LL_HUMAN STANDARD; PRT; 881 AA.
 AC Q9HAG0; O15046;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BAND 4.1-LIKE PROTEIN 1.
 GN EPB41L1 OR KIAA0338.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RT DNA Res. 4:141-150(1997).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP Tromans A.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB002336; BAA20796.1; ALT_INIT.
 DR EMBL; AL121895; CAC09920.1; -;
 DR MIM; 602879; -;
 DR InterPro; IPR000299; Band_4.1.
 DR Pfam; PF00373; Band_41; 1.
 DR PRINTS; PR00935; BAND41.
 DR SMART; SM00295; B41; 1.
 DR PROSITE; PS00660; BAND_41_1; 1.
 DR PROSITE; PS00661; BAND_41_2; 1.
 DR PROSITE; PS50057; BAND_41_3; 1.
 KW Cytoskeleton.
 FT DOMAIN 94 308 BAND 4.1-LIKE.
 FT CONFLICT 484 484 MISSING (IN REF. 2).
 FT CONFLICT 729 729 MISSING (IN REF. 2).
 SQ SEQUENCE 881 AA; 98502 MW; D923CF554EDB41D3 CRC64;
 Query Match 34.2%; Score 669.5; DB 1; Length 881;
 Best Local Similarity 41.5%; Pred. No. 8.4e-46;
 Matches 153; Conservative 59; Mismatches 138; Indels 19; Gaps 10;
 QY 18 ENSGIS---TLERGQPPPTPSKLVLS--IKIQMLDDTQEAPEVQAPGKVLDDAVCNH 72
 Db 70 EADGLSERTTPSKAQSPQKIAKKYKSAICRVTLDDASEYEVEKHGRGVFLDLVCEH 129
 QY 73 LNLVEGDYGLGLEPDPHKKITVWLDLLKPIVKQIR-RPKHVVFVVKVFPFPPDHTQLQEEL 131
 Db 130 LNLLEKDYGLTFCDAADSOKNWLDPDSKIKKQIRSSPWNFA--FTVKFYPPDPAQLTEDI 187

CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN PHOTORECEPTORS FOLLOWING
 CC THE TERMINAL MITOSIS OF RETINAL NEURONS. WHEN RETINAL
 CC SYNTOGENESIS IS COMPLETE, PROTEIN 4.1 IS ALSO EXPRESSED IN THE
 CC INNER RETINA. IN ADULT AMPHIBIAN RETINAS, PROTEIN 4.1 IS DETECTED
 CC IN PHOTORECEPTORS, BIPOLAR CELLS, AND GANGLION CELL AXONS.
 CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
 CC AND EACH PHOSPHORYLATION EVENT SELECTIVELY MODULATES THE PROTEIN'S
 CC FUNCTIONS.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M20621; AAA49695.1; --
 CC PIR: A29901; A29901.
 CC PIR: A37353; A37353.
 CC InterPro: IPR000299; Band_4.1.
 CC Pfam: PF00373; Band_41; 1.
 CC PRINTS: PR00935; BAND41.
 CC SMART: SM00295; B41; 1. 1.
 CC PROSITE: PS00660; BAND_41_1; 1.
 CC PROSITE: PS00661; BAND_41_2; 1.
 CC PROSITE: PS50057; BAND_41_3; 1.
 CC Structural protein; Cytoskeleton; Phosphorylation.
 CC KW DOMAIN 247 403 BAND 4.1-LIKE
 CC FT SEQUENCE 801 AA; 89429 MW; 07FA508552359A0F CRC64;
 CC SQ
 CC
 CC Query Match 33.6%; Score 658; DB 1; Length 801;
 CC Best Local Similarity 41.5%; Pred. No. 6.2e-45;
 CC Matches 139; Conservative 60; Mismatches 116; Indels 20; Gaps 6;
 CC
 CC QY 28 GQKPPPTSGKL-----VSIIQMLDDTQAEVPPORAGPKVLLDVCNHLNL 75
 CC DB 169 GKEPIKPEGESKASHKVVRRSPNCRKVTLLDDTVYCDLKHAKGQDIFKVCVSHLNI 228
 CC QY 76 VEGDYFGLFEPDHHKKTVMLLDLKLVKQIRPKHVVVVFV--VKFFPPDHTOLOQELTR 133
 CC DB 229 VEEDYFGLAINWSPCTKVMPLDKTRKQVHGPP---CEFTSNVKFYPPDPAQLSSEDIR 285
 CC QY 134 YLFALQVKQDLAQRCLTCDNTSAALLSHIVQSEIGDFDEALDR-EHLAKNKYIPQODA- 191
 CC DB 286 YLCLQLRXRDIFSGRLPCSFATLALLGVTVQSEVGDYEDLHGVDYVSEFKLSPNQTKD 345
 CC QY 192 LEDKIVFHHNHGQTPASDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV 251
 CC DB 346 LEEKVGLHKSYSRMTPAQADLEFLENAKKLTWYGVYDIHQAKDLEGVDIKLGVCSGLMV 405
 CC QY 252 FQGTETKINAFNAVKRLGFKRRFLKLRPDANSAYQDTLEFLMARDFCKSFWKICVE 311
 CC DB 406 FKDNLRINRPWPVKLVKISYKRSRFFIKIRPGEQEYESTIGFKLPSYKRAKKLWKCVE 465
 CC QY 312 HHAFRLFEPPKPKPKVLFVSRGSSFRSGRTQKQ 346
 CC DB 466 HHTFFRL-TSTESIPKHFRLSLGSLTFRYSGRTOAQ 499

RESULT 6

PTP1_CAEEL STANDARD; PRT; 1026 AA.
 AC P28191; P45449;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE 1 (EC 3.1.3.48).
 OS PTP-1 OR C48D5.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Lightning J., Thomas K.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 844-950 FROM N.A.
 RX MEDLINE=91139172; PubMed=1704870;
 RT "Protein tyrosine phosphatase domains from the protochordate Styela
 RT plicata.";
 RL Immunogenetics 33:33-41(1991).
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z36237; CAA85272.1; --
 CC EMBL: Z48241; CAA85272.1; JOINED.
 CC EMBL: Z48241; CAA88287.1; --
 CC EMBL: Z36237; CAA88287.1; JOINED.
 CC EMBL: M38013; AAA28127.1; --
 CC HSSP: P28827; IRPM.
 CC WormPep: C48D5.2; CEI7578.
 CC InterPro: IPR000299; Band_4.1.
 CC InterPro: IPR001478; PDZ.
 CC InterPro: IPR000387; Tyr_phosphatase.
 CC InterPro: IPR000242; Tyr_prot_phptase.
 CC Pfam: PF00373; Band_41; 1.
 CC Pfam: PF00595; PDZ; 1.
 CC Pfam: PF00102; Y_phosphatase; 1.
 CC PRINTS: PR00700; PRTYPHPTASE.
 CC PRINTS: PR00935; BAND41.
 CC SMART: SM00295; B41; 1.
 CC SMART: SM00194; PTPC; 1.
 CC PROSITE: PS00660; BAND_41_1; 1.
 CC PROSITE: PS00661; BAND_41_2; 1.
 CC PROSITE: PS50057; BAND_41_3; 1.
 CC PROSITE: PS50106; PDZ; 1.
 CC PROSITE: PS500383; TYR_PHOSPHATASE_1; 1.
 CC PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 CC PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 CC Hydrolase; Cytoskeleton.
 CC KW DOMAIN 87 245 BAND 4.1-LIKE.
 CC FT DOMAIN 777 1026 PROTEIN-TYROSINE PHOSPHATASE.
 CC FT ACT_SITE 952 952 BY SIMILARITY.
 CC SEQUENCE 1026 AA; 115093 MW; 2F43F7A614EDBC59 CRC64;
 CC SQ

Query Match 31.2%; Score 610; DB 1; Length 1026;

Best Local Similarity 39.1%; Pred. No. 6e-41;
Matches 147; Conservative 60; Mismatches 145; Indels 24; Gaps 10;

QY 13 RLGAPENSGISTLERGOKPPTPSGKLVSIKIQMLDDTQAEVPPORAGPKVLLDVCNHL 72
 DB 2 RLGSNDYVQRTAIGOTPVKTPPPNQIRCTVTFLOSTSYHFEIKNLSGLVLEKVFNY 61
 QY 73 LNLVEGDYFGLF-----PDHKITTVLWDLKLPVVKQ-IRRPKHVVVVFVFFPPDH 124
 DB 62 LEIEXDYFGLVFIADVNSAQQKK---WLDPSKLNLRKQMICPPYHLF--FRYKVVYRDP 116

[illegible]

Db 134 IAAQCAVAIAELGDHDDYKTAGVSEYRVPOQKELEGAIERIHKTLMGQAPSEAE 193
QY 213 FOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTTKINAFNNAKVRKLSFK 272
Db 194 LNYLRTAKSLEMYGVDLHPVYGENKSEYFLGLTPSGVVVYKKNQGVYFWFRITKVHF 253
QY 273 RKRFLIK-LRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRLPEEPKPKPVLF 331
Db 254 ETQFELRVLGKDCN---ETSFEEARSACKHLWKCSVEHHTTFRMPDPTESNLSRLK 310
QY 332 SRGS-SRFSGRTQKV 347
Db 311 KEGSISYKHRYTALQM 327

RESULT 12

PTNL_RAT PTNL_RAT STANDARD; PRT: 1175 AA.
AC Q62728; Q62732;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)
GN PTPN21 OR PTP2E.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE=95104449; PubMed=7805871;
RA L'Abbe D., Banville D., Tong Y., Stocco R., Masson S., Ma S.,
RA Fantus G., Shen S.H.;
RT Identification of a novel protein tyrosine phosphatase with sequence
homology to the cytoskeletal proteins of the band 4.1 family.;
RL FEBS Lett. 356:351-356(1994).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
PROTEIN TYROSINE + PHOSPHATE.
CC -!- TISSUE SPECIFICITY: PARTICULARLY ABUNDANTLY IN ADRENAL GLANDS.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17971; AAA62153.1; -;
CC EMBL; U18293; AAA62154.1; -;
CC HSSP; P18031; 1PTT.
CC InterPro; IPR000299; Band.4.1.
CC InterPro; IPR000387; TYR_phosphatase.
CC InterPro; IPR000242; Tyr_prot_phphatase.
CC Pfam; PF00373; Band.41; 1.
CC Pfam; PF00102; Y_phosphatase; 1.
CC PRINTS; PR00700; PRTYPHPHTASE.
CC PRINTS; PR00935; BAND41.
CC SMART; SM00295; B41; 1.
CC SMART; SM00194; PTPC; 1.
CC PROSITE; PS00860; BAND_41_1; 1.
CC PROSITE; PS00661; BAND_41_2; 1.
CC PROSITE; PS00507; BAND_41_3; 1.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
CC PROSITE; PS00556; TYR_PHOSPHATASE_2; 1.
KW Structural protein; Cytoskeleton; Hydrolase; Alternative splicing.

FT DOMAIN 77 241 BAND 4.1-LIKE.
FT DOMAIN 922 1175 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 1109 BY SIMILARITY.
FT VARSPLIC 1 839 MISSING (IN ISOFORM 2E).
SQ SEQUENCE 1175 AA; 133411 MW; 82A684F1C0F5ECF7 CRC64;
Query Match 18.8%; Score 368; DB 1; Length 1175;
Best Local Similarity 30.8%; Pred. No. 1.6e-21;
Matches 102; Conservative 65; Mismatches 140; Indels 24; Gaps 10;
QY 25 LERGOKPPPTSGKLVSIKIOMLDDTQEAFFVQPARPGKVLDDAVCNHNLNVEGDYFGL 84
Db 9 LKTRRYTVSSKSLVA-RIQLNNEFEVETLSVESTGQESLEAVAQRLREITVYFSLW 67
QY 85 FPDHKKITVMDLLKPIVKIRPKHV---VKFVVVFFPPDHTQLOEELRYLFAQVK 141
Db 68 YYNKQNRWRWDLKPKQL--DKHALEPTVVFVGVVPSVSQLQEQETTRYQYVQLK 125
QY 142 QDLAQRLTCNDTSRAALLSHIVQSEIGDFDEALDREHLAKNKYIP-----QDALED-- 194
Db 126 KDVLGNLPTTLEQAIQLAGLAVQADFGDFDQYESQDFLOKFAALLPVGLQDEKLEEA 185
QY 195 -KIVEFHNNHIGOTPAESDFQLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV-- 251
Db 186 QKVALLHQYRGLTAPAEAMLYMQEVRMDGGEESYPKADSGQSDISGACLDGIFVKH 245
QY 252 FOGFTKINAFNNAKVRKLSFKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFNKCIVE 311
Db 246 KNGRPV-VFRWHDIANMSHKSFPALEL--ANK--EETIQFTEDMETAKYVWRLCVA 299
QY 312 HHAFRLPEEPKPKPVLF--RGSSFRFS 340
Db 300 RHKFRYLNQCNLTQQAATLNSVRRGSSRMS 330

RESULT 13

PTNL_MOUSE PTNL_MOUSE STANDARD; PRT: 1176 AA.
AC Q62136;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE PTP-RL10).
GN PTPN21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6; TISSUE=Liver;
RX MEDLINE=95140431; PubMed=7838537;
RA Higashitsuji H., Aril S., Furutani M., Inamura M., Kaneko Y.,
RA Takenawa J., Nakayama H., Fujita J.;
RT Enhanced expression of multiple protein tyrosine phosphatases in the
regenerating mouse liver: Isolation of ppp-RL10, a novel cytoplasmic-
type phosphatase with sequence homology to cytoskeletal protein
4.1.;
RL Oncogene 10:407-414(1995).
CC -!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF GROWTH AND
DIFFERENTIATION OF LIVER CELLS.
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
PROTEIN TYROSINE + PHOSPHATE.
CC -!- TISSUE SPECIFICITY: LIVER.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; D37801; BAA07053.1; -;
DR HSP; P18031; IPTT.
DR MGD; MGI:134406; Ptpn21.
DR InterPro; IPR000299; Band 4.1.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00102; Y-phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00057; BAND_41_3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
KW Structural protein; Cytoskeleton; Hydrolase.
FT DOMAIN 77 241
FT ACT_SITE 923 1176
FT ACT_SITE 1110 1110 BY SIMILARITY.
FT DOMAIN 340 343 POLY-PRO.
FT DOMAIN 565 572
SQ SEQUENCE 1176 AA; 133490 MW; 529FBE22F1335B75 CRC64;

Query Match 18.8%; Score 368; DB 1; Length 1176;
Best Local Similarity 28.8%; Pred. No. 1.6e-21;
Matches 97; Conservative 66; Mismatches 130; Indels 44; Gaps 9;

QY 25 LERQKPPPTSGKLVSIKIMLDQAEVFPQAPGVLLDAVCNHLNMGVDFGLE 84
Db 9 LKTRRYTVSSKSLVA-RIQLNNEVFETLSVESTQESLEAVAQRLREVTYSLW 67
QY 85 FPDHKKITVLDLKPVIKVRPKHV---VKFVKKFFPDHDTQLOEELTRYLALQVK 141
Db 68 YNQNQRWVDEKPKKQL--DKHALEPTVYGVLFYVPSVQLQOEITRYOYLQK 125
QY 142 QDLAQGLTCDNTSAALLSHIVQSEIGDFDEALDRHLAKNKYIPQDALEDKIVE-- 198
Db 126 KDILEGNLPCTLEHAIQLAGLVAQADFGDFQYESQDFLOKFAFPVWLQDEKVLSEA 185
QY 199 -----FHHNIGQTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILV-- 251
Db 186 QKVALHQQYRGTLTAPAEELYMQEVERMDYGEESYPAPKDSQSDISIGACLDGIFVKH 245
QY 252 FQGTKNAPNAWAKVKLSFKRRFLKLPRDANSAYQDTLEFLMASRDFCKSFWKICVE 311
Db 246 KNGRPV-VERWHDIANMHNKSFALLEL---ANK---EETIQOTEDMETAKYVWLVCVA 299
QY 312 HHAFRL-----FEPPKPKP 326
Db 300 RHKFYRLNQCISLQATNLNVRDSSRMSLPKQP 336

RESULT 14
MOES_MOUSE STANDARD; PRT; 576 AA.
AC P26041;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MOESIN (MEMBRANE-ORGANIZING EXTENSION SPIKE PROTEIN).
GN MSN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93055012; PubMed=1429901;
RA Sato N., Funayama N., Nagafuchi A., Yonemura S., Tsukita S.,
RA Tsukita S.;
RT "A gene family consisting of ezrin, radixin and moesin. Its specific
RT localization at actin filament/plasma membrane association sites.";
RL J. Cell Sci. 103:131-143(1992).
RN [2]
RP SEQUENCE OF 11-576 FROM N.A.
RX MEDLINE=92243764; PubMed=1573844;
RA Furthmayr H., Lankes W.T., Amieva M.R.;
RT "Moesin, a new cytoskeletal protein and constituent of filopodia: its
RT role in cellular functions.";
RL Kidney Int. 41:665-670(1992).
CC -!- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
CC STRUCTURES TO THE PLASMA MEMBRANE.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is 'copyright'. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

EMBL; S47577; AAA11762.1; -;
DR EMBL; M86390; AAA39728.1; -;
DR MGD; MGI:97167; Msn.
DR InterPro; IPR000299; Band 4.1.
DR InterPro; IPR000798; Ezrin_radixin_moesin.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00769; ERW; 1.
DR PRINTS; PR00661; ERMFAMILY.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00057; BAND_41_3; 1.
KW Structural protein; Cytoskeleton.
FT INIT_MET 0 BY SIMILARITY.
FT DOMAIN 57 224
FT DOMAIN 330 331 EL -> DV (IN REF. 2).
FT CONFLICT 370 371 RA -> SP (IN REF. 2).
SQ SEQUENCE 576 AA; 67635 MW; 5E0F455552E9145 CRC64;

Query Match 18.6%; Score 363.5; DB 1; Length 576;
Best Local Similarity 27.9%; Pred. No. 1.4e-21;
Matches 98; Conservative 77; Mismatches 143; Indels 33; Gaps 10;

QY 38 KLVSIKIMLDQAEVFPQAPGVLLDAVCNHLNMGVDFGLEFPDHKKITVWLDL 97
Db 2 KTISVRTTDAELE-FAIQPTNTTQGLFDQVVKRTIGLREVWFFGLQYQDTKAFSTWLK 60
QY 98 LKPIVKQ-IRRPKHVVVKKVFPDP-HPQLOEELTRYLALQVKODLAQGLTCNDTS 155
Db 61 NKKVTADVRKESPLLFKFRKFPEDVSELIQDITQRLFFLOVKEGILNDIYCPET 120
QY 156 AALLSHIVQSEIGDFDEALDR-HLAKNKYIPQ-----QDALEKIVEFHHNHIG 205
Db 121 AVLASAVQSKYGDENKEVHKSYLAGDKLLPQRLVLEQHKLNKDQWEERIQWHEHRG 180
QY 206 QTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVQGTGKIN---AFN 262
Db 181 MLREDAVLEYIKTAQDLEMYGVNYFSITKNKKSGMLGVDALGILNIYQNDRLTPKIGFP 240
QY 263 WAKVRKLSFKRRFLI----KLPRDANSAYQDTLEFLMASRDFCKSPWKICVEHHAFRL 318
Db 241 WSEIRNISFNKKFVKIPDKKAPD-----FVFYAPRLRINKRILALCMGNHLYM- 291

```
QY 319 FEEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKKVQF-ERKHSKI 368
Db 292 ---RRRKPDITIEVQOMKAQAREKHQKQMERALLENEKKKKRLAEKEKEKI 339

RESULT 15
MOES_HUMAN
ID MOES_HUMAN STANDARD; PRT; 576 AA.
AC P26038;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MOESIN (MEMBRANE-ORGANIZING EXTENSION SPIKE PROTEIN).
GN MSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15; 53-59 AND 413-434.
RC TISSUE=Placenta;
RX MEDLINE=92020840; PubMed=1924289;
RA Lankes W.T., Furthmayr H.;
RT "Moesin: a member of the protein 4.1-talin-ezrin family of proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8297-8301(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Howden P.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
CC STRUCTURES TO THE PLASMA MEMBRANE.
CC !- TISSUE SPECIFICITY: IN ALL TISSUES AND CULTURED CELLS STUDIED.
CC !- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69066; AAA36322.1; -
CC DR EMBL; Z98946; CAB46379.1; -
CC PIR; A41289; A41289.
CC DR Aarhus/Ghent-2DPAGE; 3515; IEF.
CC DR Aarhus/Ghent-2DPAGE; 3516; IEF.
CC MIM; 309845; -
CC DR InterPro; IPR000299; Band_4.1.
CC DR InterPro; IPR000798; Ezrin_radixin_moesin.
CC Pfam; PF00373; Band_41; 1.
CC Pfam; PF00769; ERM; 1.
CC DR PRINTS; PR00661; ERMFAMILY.
CC DR PRINTS; PR00935; BAND41.
CC DR SMART; SM00295; B41; 1.
CC DR PROSITE; PS00660; BAND_41_1; 1.
CC DR PROSITE; PS00661; BAND_41_2; 1.
CC DR PROSITE; PS00507; BAND_41_3; 1.
CC KW Structural protein; Cytoskeleton.
CC INIT MET 0
FT DOMAIN 57 224 BAND 4.1-LIKE.
FT SEQUENCE 576 AA; 67689 MW; 96CE7C28D55A8176 CRC64;
```

```
Query Match 18.5%; Score 362.5; DB 1; Length 576;
Best Local Similarity 27.9%; Pred No. 1.7e-21;
Matches 98; Conservative 77; Mismatches 143; Indels 33; Gaps 10;

QY 38 KLVSTIKQLDDTQEAFAFVQAPRGKVLDAVCNHLNLVVDGDFGLFDPDHKKITVWLDL 97
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2 KTISVRVTTMDAELE-FAIQPNNTGKQLFDQVVKVTIGLREVWFFGLQYQDTKGFSWLKL 60

QY 98 LRPPIVKQ-IRRPKHVVVKKFVFFPPD-HTQLQEELTRYLFALQVKQDLAQGRLCNDTS 155
```

```
Db 61 NKKVTAQDVKRESPLLKFRAFYEDVSEELIQDITQRLFFLQVKEGILNDDIYCPTT 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 156 AALLISHIVQSEIGDFDEALDRE-HLAKNKYIPQ-----QDALEDKIYVEFHHNHIG 205
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 121 AVLLASYAVQSKYGVDFNKVKHSGYLAGDKLLPQRVLEQHKLNKQDQWEERIQWHEEHG 180
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 206 QTPASDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILYVFGFTKIN---AFN 262
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 MLREDAVLEYLKIAODLEMYGVNYFSIKNKGSELWGLVDALGLNIYEQNDRLTPKIGFP 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 263 WAKVRKLSEFKRRFLI----KLRPDANSAYQDTLEFLMASRDFCKSFWRICVVEHHAFRL 318
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 241 WSEIRNISFNKKFVYKPIDKKAPD-----FVYAPRLRINKRILALCMGNHLYM- 291
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 319 FEEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKKVQF-ERKHSKI 368
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 292 ---RRRKPDITIEVQOMKAQAREKHQKQMERALLENEKKKKREMAEKEKEKI 339
```

Search completed: December 6, 2001, 08:53:08
Job time: 380 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:54:30 ; Search time 77.11 Seconds
(without alignments)
709.452 Million cell updates/sec

Title: US-09-555-342A-2_COPY_1_374

Perfect score: 1956

Sequence: 1 MGTEQRPTGSLGAPENS.....GHKKVQPERKXKIHRSRL 374

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTEMBL_17.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-virus.*
- 13: sp-vertebrate.*
- 14: sp-unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	1956	100.0	1045	Q9Y4F1	Q9y4f1 homo sapien
2	1246.5	63.7	1054	Q94887	Q94887 homo sapien
3	867	44.3	330	Q9VN68	Q9vn68 drosophila
4	802.5	41.0	1111	Q17905	O17905 caenorhabdi
5	699.5	35.8	988	11 070318	O70318 mus musculu
6	691.5	35.4	929	11 Q9WV92	Q9wv92 mus musculu
7	691.5	35.4	963	11 Q9JMB2	Q9jmb2 rattus norv
8	691.5	35.4	1105	11 Q9JMB3	Q9jmb3 rattus norv
9	690.5	35.3	865	4 Q9BRP5	Q9brp5 homo sapien
10	690.5	35.3	1115	4 Q9Y2J2	Q9y2j2 homo sapien
11	689.5	35.3	1005	4 Q43491	O43491 homo sapien
12	684	35.0	503	4 Q95713	Q95713 homo sapien
13	681.5	34.8	864	4 Q9T579	Q9t579 homo sapien
14	677	34.6	706	11 Q9R102	Q9r102 mus musculu
15	671.5	34.3	617	6 Q9N179	Q9n179 bos taurus
16	670.5	34.3	641	4 Q14245	Q14245 homo sapien
17	669.5	34.2	879	4 Q9H4G0	Q9h4g0 homo sapien
18	669.5	34.2	934	4 Q15046	O15046 homo sapien
19	667.5	34.1	879	11 Q9Z2H5	Q9z2h5 mus musculu

Query Match 100.0%; Score 1956; DB 4; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.6e-148;

20	666.5	34.1	879	11 Q9WTP0	Q9wtp0 rattus norv
21	666.5	34.1	1551	11 Q9WTP1	Q9wtp1 rattus norv
22	662.5	33.9	926	11 Q9WU22	Q9wu22 mus musculu
23	661.5	33.8	703	5 Q9V8S0	Q9v8s0 drosophila
24	661.5	33.8	889	5 Q9V8R8	Q9v8r8 drosophila
25	661.5	33.8	1698	5 Q24440	Q24440 drosophila
26	661.5	33.8	1698	5 Q9V8R9	Q9v8r9 drosophila
27	589.5	30.1	505	6 Q9WU08	Q9wyu8 canis famil
28	587.5	29.9	687	4 Q9H975	Q9h975 homo sapien
29	584.5	29.9	504	11 Q99K28	Q99k28 mus musculu
30	567	29.0	972	5 Q9VFU8	Q9vf8 mus musculu
31	552.5	28.2	527	11 Q9JMC8	Q9jmc8 mus musculu
32	542.5	27.7	504	4 Q9H328	Q9h328 homo sapien
33	542.5	27.7	518	4 Q9P2V3	Q9p2v3 homo sapien
34	542.5	27.7	913	4 Q9H329	Q9h329 homo sapien
35	531.5	27.2	987	5 Q9NDP4	Q9ndp4 ciona intes
36	527	26.9	568	5 Q9GP93	Q9gp93 caenorhabdi
37	524	26.8	283	4 Q9BQV2	Q9bv2 homo sapien
38	514.5	26.3	440	4 Q9NX84	Q9nx84 homo sapien
39	480.5	24.6	598	4 Q9HCS5	Q9hcs5 homo sapien
40	438.5	22.4	916	5 Q9W0R3	Q9w0r3 drosophila
41	434.5	22.2	1199	5 Q9W9Y3	Q9w9y3 drosophila
42	423	21.6	639	5 Q9XV97	Q9xv97 caenorhabdi
43	423	21.6	4667	5 Q9TVI9	Q9tvi9 caenorhabdi
44	416.5	21.3	572	5 Q9VKY7	Q9vky7 drosophila
45	363.5	18.6	577	11 Q35763	O35763 rattus norv

ALIGNMENTS

RESULT 1

Q9Y4F1 PRELIMINARY; PRT: 1045 AA.
AC Q9Y4F1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDEP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CARTILAGE;

RX MEDLINE=98086358; PubMed=9425278;

RA Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,

RA kato Y.;

RT *Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors.*;
RL Biochem. Biophys. Res. Commun. 241:369-375(1997).

DR EMBL; AB008430; BAA24267.1; .

DR HSSP; P08567; IPLS.

DR InterPro; IPR000299; Band_4.1.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000219; RhoGEF.

DR Pfam; PF00373; Band_41; 1.

DR Pfam; PF00169; PH; 2.

DR Pfam; PF00621; RhoGEF; 1.

DR PRINTS; PR00935; BAND41.

DR SMART; SM00295; B41; 1.

DR SMART; SM00233; PH; 2.

DR SMART; SM00325; RhoGEF; 1.

DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.

DR PROSITE; PS50057; BAND_41_3; 1.

DR PROSITE; PS50003; PH_DOMAIN; 2.

SQ SEQUENCE 1045 AA; 118632 MW; 0F8B2D61C0F58417 CRC64;

Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGEIEQRTGSRGAPENSIGSTLERGQKPPPTSGKLVSIKIOMLDDTQEAFAEPQRA 60
 DB 1 MGEIEQRTGSRGAPENSIGSTLERGQKPPPTSGKLVSIKIOMLDDTQEAFAEPQRA 60

QY 61 PGKVLDAVNCNHLNVEGDYFGLFDPHKKITVWLDLLKPIVQIRRPKHVVVKKFF 120
 DB 61 PGKVLDAVNCNHLNVEGDYFGLFDPHKKITVWLDLLKPIVQIRRPKHVVVKKFF 120

QY 121 PPQHTQLEELTRYLFALQVQDLAOGRLTCNDTSAALLSHIVQSEIGDFDEALDREHL 180
 DB 121 PPQHTQLEELTRYLFALQVQDLAOGRLTCNDTSAALLSHIVQSEIGDFDEALDREHL 180

QY 181 AKNKYIPQDALEDKIVEFHNNHIGOTPAESDFQLEIARLEMYGIRLHPAKDREGTKI 240
 DB 181 AKNKYIPQDALEDKIVEFHNNHIGOTPAESDFQLEIARLEMYGIRLHPAKDREGTKI 240

QY 241 NLAVANTGILVFGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRD 300
 DB 241 NLAVANTGILVFGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRD 300

QY 301 FCKSFWKICVEHHAFFLFEPEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKKVQ 360
 DB 301 FCKSFWKICVEHHAFFLFEPEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKKVQ 360

QY 361 FERKSHKHSIRSL 374
 DB 361 FERKSHKHSIRSL 374

RESULT 2

ID O94887 PRELIMINARY; PRT; 1054 AA.

AC O94887;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE KIA00793 PROTEIN.

GN KIA00793.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=99087487; PubMed=9872452;

RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RA "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RT DNA Res. 5:277-286(1998).

RL [2]

RN SEQUENCE OF 1-754 FROM N.A.

RP MEDLINE=99063792; PubMed=9847074;

RA Sulston J.E., Waterston R.;

RT "Toward a complete human genome sequence.";

RL Genome Res. 8:1097-1108(1998).

RN [3]

RN SEQUENCE OF 1-754 FROM N.A.

RA Sun H., Stoneking T., Langston Y., LaPlant Y.;

RT "The sequence of Homo sapiens BAC clone RC442F18.";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RN SEQUENCE OF 1-754 FROM N.A.

RA Waterston R.H.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [5]

RN SEQUENCE OF 1-754 FROM N.A.

RA Waterston R.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

[6]

RN SEQUENCE OF 1-754 FROM N.A.

RP Waterston R.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB018336; BAA34513.1; -

DR EMBL; AC005104; AAD12224.1; -

DR HSSP; P08567; IPLS.

DR InterPro; IPR000299; Band_4.1.1.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000219; RhoGEF.

DR Pfam; PF00373; Band_41; 1.

DR Pfam; PF00169; PH; 2.

DR Pfam; PF00621; RhoGEF; 1.

DR PRINTS; PR00935; BAND41.

DR SMART; SM00295; B41; 1.

DR SMART; SM00233; PH; 2.

DR SMART; SM00325; RhoGEF; 1.

DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.

DR PROSITE; PS00057; BAND_41_3; 1.

DR PROSITE; PS00003; PH_DOMAIN; 2.

SQ SEQUENCE 1054 AA; 119887 MW; 7DC279F69A307E5A CRC64;

Query Match 63.7%; Score 1246.5; DB 4; Length 1054;
 Best Local Similarity 62.5%; Pred. No. 1.5e-91;
 Matches 237; Conservative 58; Mismatches 79; Indels 5; Gaps 2;

QY 1 MGEIEQR---PTPGSRGAPENSIGSTLERGQKPPPTSGKLVSIKIOMLDDTQEAFAEV 56
 DB 1 MGEIEGTYRVLTAGNRLGAQTPGVSTLPGQTLPRMQEKLHLKLVKLLDNTMEIFDI 60

QY 57 PQRAPGKVLDAVNCNHLNVEGDYFGLFDPHKKITVWLDLLKPIVQIRRPKHVVVKKV 116
 DB 61 EPKCDGQVLLTQVWKRLNLVECDYFGMEFQNTQSYWLEPKPIIRQIRRPKNVVLRLA 120

QY 117 VKFFPDHTQLQEELTRYLFALQVQDLAOGRLTCNDTSAALLSHIVQSEIGDFDEALD 176
 DB 121 VKFFPDHTQLQEELTRYLFALQVQDLAOGRLTCNDTSAALLSHIVQSEIGDFDEALD 180

QY 177 REHLAKNKYIPQDALEDKIVEFHNNHIGOTPAESDFQLEIARLEMYGIRLHPAKDRE 236
 DB 181 REHLKVNELVPGQOHCLEKLEFHQKVGOTPAESDFQLEIARLEMYGIRLHPAKDRE 240

QY 237 GTKINLAVANTGILVFGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLM 296
 DB 241 GTKIQAVSHMGVLVFGFTKINTFNWSKVRKLSFKRKRFLIKLRHPVHGPGYQDTLEFL 300

QY 297 ASRDFCKSWKICVEHHAFFLFEPEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGH 356
 DB 301 GSRDECKNFWKICVEYTFEFLDQPKAKAVFFSGSSFRSGRTQKQVLDYVKEGGH 360

QY 357 KKVQFERKSHKHSIRSL 374
 DB 361 KRIPIERRHKSHTSVAL 379

RESULT 3

QYVN68 PRELIMINARY; PRT; 330 AA.

AC QYVN68;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE CG1283 PROTEIN.

GN CG1283.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K., Gabrieli A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195 (2000).
DR EMBL; AE003604; AAF52081.1;
DR FlyBase; Fg0037286; CG1283.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS0057; BAND_41_3; 1.
SQ SEQUENCE 330 AA; 38707 MW; FBDAB30DBC86CBF CRC64;

Query Match 44.3%; Score 867; DB 5; Length 330;
Best Local Similarity 51.5%; Pred. No. 9.2e-62;
Matches 168; Conservative 64; Mismatches 84; Indels 10; Gaps 6;

QY 46 MLDSTQAEVQAPRGVLLDAVCHNLVSGDFEGLEFPD---HKITVWLDLKPV 102
DB 1 MLDSTMTFQVQAKALGRVLFQVQCQLNLEADYFGLYQEVSTHTK--YWLDEKPNW 58
QY 103 KQLRRPK-HVVYKFFPPDHTQLQELRYLFLALQVKQDLAOGRLTCNDTSAALLS 161
DB 59 RQVGLSLDPVLCFKFTTPDPAQLLEEYTRYLFCQLKRDLATGSLCNDNTAALMAS 118
QY 162 HIYQSEIGDF--DEALDREHLAKNKYIPOQDA-LEDKIVEFHNNHIGQTPAESDFOLLEI 218
DB 119 YIVQASCGDFVPDYPDHTYLSYRPVQDATMQRKIMENHKHKGVSQPAEADNLLET 178
QY 219 ARLEMYGRLHPAKDREGTKINLAVANTGILVFGQFTKINAFNWKVKRLSKFRKRLI 278
DB 179 ARCELYGKMHKPAKDVGVPLNLAHMGITVFQNTIRNTFSWAKIRKISFKRKRFLV 238
QY 279 KLRPDANSAYQDTLEFLMASRDFCKSFVKICVHHAFRLFE-EPKPKPKVPLFSRGSF 337
DB 239 KLPHEGYGYKDTVEFFFGRECKNFWKCKVENHGFRCCTAVQNTPRKTRVLSRGSF 298

QY 338 RFSGRTQKOVLDYVKEGGHKKVQFER 363
DB 299 RYSGRTQKQIEFVRENVYKRONFOR 324
RESULT 4
O17905 PRELIMINARY; PRT; 1111 AA.
AC O17905;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE H05G16.1 PROTEIN.
GN H05G16.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smailson N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
RL Nature 368:32-38 (1994).
DR EMBL; Z97190; CAB10024.1;
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS0057; BAND_41_3; 1.
DR SMART; SM00295; B41; 1.
SQ SEQUENCE 1111 AA; 124808 MW; 1EE53EB512C0FE2A CRC64;

Query Match 41.0%; Score 802.5; DB 5; Length 1111;
Best Local Similarity 44.2%; Pred. No. 7.1e-56;
Matches 165; Conservative 64; Mismatches 113; Indels 31; Gaps 6;

QY 22 ISTLERGQKPPTPS-----GKLSIKIOMLDDTQEAPE----- 55
DB 1 MSNIPRGVAGPPPGMNSKRGRLVCIKVMMLDDTVAVFHLGVICFFCVSWPHGYVFL 60
QY 56 VQRAPGKVLDDAVCHNLVSGDFEGLEFPDHTKKTIVWLDLKPVKQIRR-PKHVVYK 114
DB 61 LQKALGQTLLEDCVCHNLLECDYGLSFIDINGNHCWLDREKTLRQITNGSTDAKY 120
QY 115 FVYKFPDHTQLQELRYLFLALQVKQDLAOGRLTCNDTSAALLSHVQSEIGDF--D 172
DB 121 FVYKFTPNPDLDEEYTRYLFTMQIKRDLALGELCHSDNTASLLSAYLVQSCGDSSE 180
QY 173 EALDREHLAKNKYIPOOD-ALEDKIVEFHNNHIGQTPAESDFOLLEIARRLEMYGRLHP 231
DB 181 DYPDATYLSHTFRVNPQTLFQKKVMNDNRNFTGMTPGESDLAMLEVARCQFYGVKLHA 240
QY 232 AKDREGTKINLAVANTGILVFGQFTKINAFNWKVKRLSKFRKRLIKLRPDANSAYQDT 291
DB 241 AKDIDGNDALSVNHLGIKVFQRLQDLDTFSWARIRKLSFKRKKLVKLPDPSQYLKET 300

```
QY 292 LEFTMASRDCCKSWKICVHHAFRLFEPPKPKPKVLF-SRGSSFRSGRTQKQVLDY 350
Db 301 VEFSETRDECKNFKWCKVCHHAFRCVQAEPPKPKKTRFFISKSSPRYHGRTKQLIDY 360
QY 351 VKEGGHKKVOFER 363
Db 361 VREHHKREPPTR 373

RESULT 5
O70318 PRELIMINARY; PRT; 988 AA.
AC O70318;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTEIN 4.1G (FRAGMENT).
GN EPB4.1L2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98198473; PubMed=9531554;
RA Walensky L.D., Gascard P., Fields M.E., Blackshaw S., Conboy J.G.,
RA Mohandas N., Snyder S.H.;
RT "The 13-kD FK506 binding protein, FKBP13, interacts with a novel
RT homologue of the erythrocyte membrane cytoskeletal protein 4.1.";
RL J. Cell Biol. 141:143-153(1998).
DR EMBL; AF044312; AAC40083.1;
DR MGD; MGI:103009; Epb4.1L2.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_4.1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
FT NON_TER 988
SQ SEQUENCE 988 AA; 109833 MW; B0367A16C5A2EC05 CRC64;

Query Match 35.8%; Score 699.5; DB 11; Length 988;
Best Local Similarity 43.7%; Pred. No. 1.1e-47;
Matches 150; Conservative 56; Mismatches 132; Indels 5; Gaps 5;

QY 34 TPGSKLVSIKIQMLDQTOEAFVQAPGKVLDDAVCNHLNLVEGDFGLFPPDHKKITV 93
Db 205 TKTKTKVLAKVTLLDGTETSCDLEKRAKGQVDFRVCHEHLNLEKDYFGLLFQDHPEQN 264
QY 94 WLDLKPIVKQIRRRKHVVVKKFPPDPDHTQQLQELTRYLFALQVKODLAQGRLTCD 153
Db 265 WLDPKAEIKQLNKLWLF-FNVKYPDPDSQLTEDITRYFLCLQRLQDIASGRPCSF 323
QY 154 TSAALLISHIVQSEIGDFD-EALDREHLAKNKYIP-QQDALEDKIVEFHNNHIGOTPAES 211
Db 324 VTALLGSYTLQAEHGDPDEYDSIDLQDFQAPAHTEKLEKVSSELKHTHRLSPAQA 383
QY 212 DFQLELTARLEMYGTLHPAKDREGTKINLAVANTGILVFGQFTKINAFNWKVKLSF 271
Db 384 DSQFLENAKRLSMYGVLDLHAKDSEGVDIKLGVCANGLLIYKDRLRINRFAMPKILKISY 443
QY 272 KRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVHHAFRLFEPPKPKPVLF 331
Db 444 KRSNFYIKVPALEQEFESTIGFKLPHNRAAKRLKVCVHHHTFYRLV-SPEQPPKTKFL 502
QY 332 SRGSSFRSGRTQKQVLDYVKEGGHKKVOFERKHSIHRSLS 374
Db 503 TLGSKFRYSGRTQATREASTLIDRAPQOFERASSKRVS-RSL 544

RESULT 6
```

Q9WV92

```
ID Q9WV92 PRELIMINARY; PRT; 929 AA.
AC Q9WV92;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTEIN 4.1B.
GN EPB4.1L3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Parra M.K., Chan N.L., Gascard P.D., Walensky L.D., Lee G.L.,
RA Chasis J.A., Snyder S.H., Narla M., Conboy J.G.;
RT "Characterization of protein 4.1B, a new gene in the protein 4.1
RT family with high level, focal expression in brain.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152247; AAD38048.1;
DR MGD; MGI:103008; Epb4.1L3.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
SQ SEQUENCE 929 AA; 103337 MW; F4975FF405DA44AE CRC64;

Query Match 35.4%; Score 691.5; DB 11; Length 929;
Best Local Similarity 41.9%; Pred. No. 4.5e-47;
Matches 153; Conservative 62; Mismatches 133; Indels 17; Gaps 9;

QY 18 ENSGISTLREQQ---KPPPTSGKLVSIKIQMLDQTOEAFVQAPGKVLDDAVCNHL 73
Db 97 QRSSSSKLSRSLKIVKRP-----KSMQCKVTLLDGSSEYGCVDKSRGQVLFDKVCEHL 151
QY 74 NLVEGDFGLFPPDHKKITVWLDLKPVIKQIRRRPK-HVVVKKFVVKFFPDHTQQLQELT 132
Db 152 NLEKDYFGLTYRDAENQKNWLDPAKEIKKQIRSGAWH--FSFNVKFPYPPDPAQLSEDT 209
QY 133 RYLFALQVKODLAQGRLTCDNTSAALLISHIVQSEIGDFD-EALDREHLAKNKYIPQ-D 190
Db 210 RYLCQLQRRDQIVSGRLPCSFVTLALLGSYTVQSELGDYDPECGNDYISERFANHTK 269
QY 191 ALEDKIVEFHNNHIGOTPAESQQLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 270 ELEDKVIELHKSHRGMPAEAEHMFLENAKKLSMYGVLDLHAKDSGVETMLGVCASGLL 329
QY 251 VFGQFTKINAFNWKVKLSFRKKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICV 310
Db 330 IYDRLRINRFAMPKVLKISYKRNNFYIKIRPGEFFQFESTIGFKLPHNRAAKRLKVCV 389
QY 311 EHHAFLRPEEPKPKPVLFSGRSSFRSGRTQKQVLDYVKEGGHKKVOFERKHSIHRS 370
Db 390 EHHFTFRL-LPEAPPKFL-TLGSKFRYSGRTQATRRASALIDRPAPYFERSSSKRYT 447
QY 371 I-RSL 374
Db 448 MSRSLS 452

RESULT 7
Q9JMB2 PRELIMINARY; PRT; 963 AA.
AC Q9JMB2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1 MINOR ISOFORM.
```


OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=20267856; PubMed=10806359;
RA Yamakawa H., Ohara O.;
RT "Comparison of mRNA and protein levels of four members of the protein
RT 4.1 family: The type II brain 4.1/4.1b/KIAA0987 is the most
RT predominant member of the protein 4.1 family in rat brain.";
RL Gene 248:137-145(2000).
DR EMBL; AB032827; BAA90774.1; -.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PRO0935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00557; BAND_41_3; 1.
SQ SEQUENCE 963 AA; 107071 MW; B37C137B04305F71 CRC64;

Query Match 35.4%; Score 691.5; DB 11; Length 963;
Best Local Similarity 41.9%; Pred. No. 4.8e-47;
Matches 153; Conservative 62; Mismatches 133; Indels 17; Gaps 9;

QY 18 ENSGISTLGRGQ---KPPPTPSGKLVSIKIOMLDDTQEAFAFVQAPQKGVLLDAYCNHL 73
DB 92 QKSSSKLSRSLKIVKRP-----KSMQCKVTLDDGSEYGCDDVKRSRGQVLFDKVCEHL 146

QY 74 NLVEGDYFGLPEPDHKKITVWLDLKPVIKQIRRPK-HVVVKFVVFPPDPHTQLOEELT 132
DB 147 NLEKDYFGLTYRDAENQKNWLDPAKEIKKQIRSGAWH--FSENVKVFPPDPAQLSEDI 204

QY 133 RYLFALQVKQDLAQGLTCDNTSAALLSHIVQSEIGDFD-EALDREHLAKNKYIPQQ-D 190
DB 205 RYLCQLRDDIVSGRLPCSFVTLALLGSYTVQSELGDDPDCEGNDYISEFRFAPNHTK 264

QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 265 ELEDKVIELHKSHRGTAPAEAMHFLNNAKLSMYGVDLHAKDSEGEVIMLGVCASGLL 324

QY 251 VFQGTFTKINAFNAWKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICV 310
DB 325 IYRDLRLINRFAMPKVLKISYKRNPFYIKRPGCFEFGFESTIGFKLPNHRAAKRLWKVCV 384

QY 311 EHHAFRLFEKPKPKPVLFSGRSFSGRTQKQVLDYVKGEGHKVKQVFERKHKSHS 370
DB 385 EHHTFFRL-LPEAPPKKFL-TLGSFRYSGRTOAOTRRASALIDRAPYFFERSKRYT 442

QY 371 I-RSL 374
DB 443 MSRL 447

RESULT 8
Q9JMB3 PRELIMINARY; PRT; 1105 AA.
AC Q9JMB3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=20267856; PubMed=10806359;

RA Yamakawa H., Ohara O.;
RT "Comparison of mRNA and protein levels of four members of the protein
RT 4.1 family: The type II brain 4.1/4.1b/KIAA0987 is the most
RT predominant member of the protein 4.1 family in rat brain.";
RL Gene 248:137-145(2000).
DR EMBL; AB032827; BAA90774.1; -.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PRO0935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00557; BAND_41_3; 1.
SQ SEQUENCE 1105 AA; 122451 MW; 94F1CA93A83693C6 CRC64;

Query Match 35.4%; Score 691.5; DB 11; Length 1105;
Best Local Similarity 41.9%; Pred. No. 5.7e-47;
Matches 153; Conservative 62; Mismatches 133; Indels 17; Gaps 9;

QY 18 ENSGISTLGRGQ---KPPPTPSGKLVSIKIOMLDDTQEAFAFVQAPQKGVLLDAYCNHL 73
DB 92 QKSSSKLSRSLKIVKRP-----KSMQCKVTLDDGSEYGCDDVKRSRGQVLFDKVCEHL 146

QY 74 NLVEGDYFGLPEPDHKKITVWLDLKPVIKQIRRPK-HVVVKFVVFPPDPHTQLOEELT 132
DB 147 NLEKDYFGLTYRDAENQKNWLDPAKEIKKQIRSGAWH--FSENVKVFPPDPAQLSEDI 204

QY 133 RYLFALQVKQDLAQGLTCDNTSAALLSHIVQSEIGDFD-EALDREHLAKNKYIPQQ-D 190
DB 205 RYLCQLRDDIVSGRLPCSFVTLALLGSYTVQSELGDDPDCEGNDYISEFRFAPNHTK 264

QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 265 ELEDKVIELHKSHRGTAPAEAMHFLNNAKLSMYGVDLHAKDSEGEVIMLGVCASGLL 324

QY 251 VFQGTFTKINAFNAWKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICV 310
DB 325 IYRDLRLINRFAMPKVLKISYKRNPFYIKRPGCFEFGFESTIGFKLPNHRAAKRLWKVCV 384

QY 311 EHHAFRLFEKPKPKPVLFSGRSFSGRTQKQVLDYVKGEGHKVKQVFERKHKSHS 370
DB 385 EHHTFFRL-LPEAPPKKFL-TLGSFRYSGRTOAOTRRASALIDRAPYFFERSKRYT 442

QY 371 I-RSL 374
DB 443 MSRL 447

RESULT 9
Q9BRP5 PRELIMINARY; PRT; 865 AA.
AC Q9BRP5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SIMILAR TO ERYTHROCYTE PROTEIN BAND 4.1-LIKE 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006141; RAH06141.1; -.
SQ SEQUENCE 865 AA; 96513 MW; C2070B01BF79422 CRC64;

Query Match 35.3%; Score 690.5; DB 4; Length 865;
Best Local Similarity 41.5%; Pred. No. 5e-47;
Matches 152; Conservative 64; Mismatches 131; Indels 19; Gaps 9;

OS Mus musculus (Mouse).

2. entusiasmo (noun):

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:50:18 ; Search time 36.72 Seconds
(without alignments)
640.412 Million cell updates/sec

Title: US-09-555-342A-2
Perfect score: 5463
Sequence: 1 MGIERPPTGSLGAPENS.....SATSSASRPHVLKSHKSLVY 1045

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	646	11.8	345	US-08-446-345-40	Sequence 40, Appl
2	612	11.2	288	US-09-100-804-14	Sequence 14, Appl
3	522	9.6	247	US-09-100-804-13	Sequence 13, Appl
4	500.5	9.2	288	US-09-100-804-15	Sequence 15, Appl
5	499	9.1	342	US-08-446-345-39	Sequence 39, Appl
6	396	7.2	1174	US-08-446-345-36	Sequence 36, Appl
7	344	6.3	296	US-09-100-804-12	Sequence 12, Appl
8	335	6.1	591	US-08-179-738-10	Sequence 10, Appl
9	335	6.1	591	US-08-628-145-10	Sequence 10, Appl
10	335	6.1	595	US-08-171-718-16	Sequence 16, Appl
11	335	6.1	595	US-08-478-087-16	Sequence 16, Appl
12	335	6.1	596	US-08-179-738-3	Sequence 3, Appl
13	335	6.1	596	US-08-628-145-3	Sequence 3, Appl
14	332	6.1	584	US-08-179-738-7	Sequence 7, Appl
15	332	6.1	584	US-08-628-145-7	Sequence 7, Appl
16	330	6.0	591	US-08-179-738-5	Sequence 5, Appl
17	330	6.0	591	US-08-628-145-5	Sequence 5, Appl
18	330	6.0	596	US-08-179-738-2	Sequence 2, Appl
19	330	6.0	596	US-08-628-145-2	Sequence 2, Appl
20	308	5.6	2485	US-09-290-640-46	Sequence 46, Appl
21	300.5	5.5	2466	US-09-080-855-12	Sequence 12, Appl
22	300.5	5.5	2466	PCT-US94-09943-2	Sequence 2, Appl
23	296	5.4	2465	US-08-596-291-3	Sequence 3, Appl
24	296	5.4	2465	US-09-100-804-3	Sequence 3, Appl
25	265	4.9	303	US-09-100-804-11	Sequence 11, Appl
26	264.5	4.8	2860	US-08-826-267-2	Sequence 2, Appl
27	212.5	3.9	1461	US-08-993-228-10	Sequence 10, Appl

28 206 3.8 1244 3 US-09-356-952-7 Sequence 7, Appl
29 180 3.3 761 4 US-09-625-188-14 Sequence 14, Appl
30 179.5 3.3 912 4 US-08-943-768-2 Sequence 2, Appl
31 159.5 2.9 477 1 US-08-191-338A-2 Sequence 2, Appl
32 142 2.6 1319 2 US-08-290-731C-2 Sequence 2, Appl
33 142 2.6 1336 2 US-08-290-731C-6 Sequence 6, Appl
34 140 2.6 1297 2 US-08-290-731C-4 Sequence 4, Appl
35 137.5 2.5 2409 6 5180808-2 Patent No. 5180808
36 136 2.5 1333 3 US-09-356-952-2 Sequence 2, Appl
37 136 2.5 1805 1 US-07-853-913-2 Sequence 2, Appl
38 135.5 2.5 750 4 US-08-202-841A-2 Sequence 2, Appl
39 133.5 2.4 1354 3 US-08-685-871-2 Sequence 4, Appl
40 133 2.4 620 1 US-08-484-105-4 Sequence 4, Appl
41 133 2.4 620 1 US-08-484-106-4 Sequence 4, Appl
42 128 2.3 1572 2 US-08-290-731C-5 Sequence 5, Appl
43 128 2.3 1596 3 US-09-356-952-3 Sequence 3, Appl
44 126 2.3 772 1 US-08-524-757-12 Sequence 12, Appl
45 122.5 2.2 399 4 US-09-284-033-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-446-345-40
; Sequence 40, Application US/08446345
; Patent No. 5831009
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-446-345-40

Query Match 11.8%; Score 646; DB 2; Length 345;
Best Local Similarity 40.9%; Pred. No. 9e-47;

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446, 345
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234, 440
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 7683-054
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-446-345-39

Query Match 9.1%; Score 499; DB 2; Length 342;
Best Local Similarity 34.7%; Pred. No. 2.9e-34;
Matches 114; Conservative 66; Mismatches 133; Indels 16; Gaps 7;

QY 27 RGQKPPPTSGKLVSIKIQMLDDTQEAPEVQAPQAPGVLLDVCNHLNVEGDYFGLFP 86
Db 16 RTSLEPKETSEVICSIHFLDGVQVTKQDTGGVLLDMVHNLGVTEKEIFGLQHD 75
QY 87 DHKKITV-WLDLKLPIVKQIRRPKHVVVVKVFFPPDHTQLQBELTRYLFALQVKODLA 145
Db 76 DSDVSPRWLEASPIRKQLKGGFPCTLHFRVRFIPDPNTLQEQTRHLYFLQKMDIC 135
QY 146 QGRLTCDNTSALLISHIVQSEIGDDEALDRE-HLAKNKYIQQDALED---KIVEFHH 201
Db 136 EGRUTCPNLASVAVASVAVOSVHFGDYNSTHHPGYSDSHFIPDN--EDFLTKVESLHE 193
QY 202 NHIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTKINAF 261
Db 194 QHSLKQSEAESEYINARTLDFYGVHLSGRDLHNLDMIGASAGVAYRYICTSFY 253
QY 262 NNAKVRKLSFKRRFLIKLPDANSAYQDTLEFLMASRDFCKSWKICVEHHAFFRLFE 321
Db 254 PWNILKISFKRKKFFTHQRQKQAESREHIVAFNMLNRYRCKNLWKSCVBEHHTFQA-KK 312
QY 322 PKPKPKVL---PSRGSSFRSGRTQKV 347
Db 313 LLFQEKVLSQYWTMGSR-----NTRKSV 336

RESULT 6
US-08-446-345-36
Sequence 36, Application US/08446345
Patent No. 5831009
GENERAL INFORMATION:
APPLICANT: Ullrich, Axel
APPLICANT: Moller, Niels P.H.
APPLICANT: Moller, Karin B.
TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
TITLE OF INVENTION: PHOSPHATASES PTP-DI
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446, 345
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234, 440
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 7683-054
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 1174 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-446-345-36

Query Match 7.2%; Score 396; DB 2; Length 1174;
Best Local Similarity 22.5%; Pred. No. 1.3e-24;
Matches 141; Conservative 103; Mismatches 248; Indels 134; Gaps 18;

QY 25 LERQKPPPTSGKLVSIKIQMLDDTQEAPEVQAPQAPGVLLDVCNHLNVEGDYFGL 84
Db 9 LKTRTYVSSKCLVA-RIOLLNNEFEETLSVESTGOESLEAVAQRLRELVYFSLW 67
QY 85 FPDHKITVWLDLKLPIVKQIRRPKHVVVVKVFFPPDHTQLQBELTRYLFALQVKOD 143
Db 68 YNKNQRRWVLDLEKPKQLDKYALEPTVYFGVWFVPSVLSQLOEQEITRYQLQLK 127
QY 144 LAQGRLTCDNTSALLISHIVQSEIGDDEALDREHLAKNKYIP---QQDALED---K 195
Db 128 ILEGSTPCTLEQAIQIAGLAVQADFGDQYESQDFLQKFAFPVGVWLQDEKVKLEATOK 187
QY 196 IVEFHHNHIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILV-FQG 254
Db 188 VALLHQYKRGLTAPDAEMLYMQVERMDGYGESYPAKDSQGSISIGACLEGIFVKHN 247
QY 255 FTKINAFNAKVRKLSFKRRFLIKLPDANSAYQDTLEFLMASRDFCKSWKICVEHHA 314
Db 248 GRHPVVRWHDIANMGNHNSFFALEL---ANK--EETIQFQEDMETAKYIWRCLVARHK 302
QY 315 EFRLEP-----EPKPKPKVLFSGSSFRPSGR-----TOKVL 348
Db 303 FYRLNQCNIQTQTVTNPTRRSSRSKSLPKQPYVMPPPPQLHYNGHYTEPYASSQDNL 362
QY 349 DYVKEGG---HKVQFER-----KHSKHSIKS-----LASOPTEL 381
Db 363 FVPNQNGYYCHSOTSILDRADIFNGRIRNGSVSAHSTNSLNNPQYLPQSPMSSNPSIT 422
QY 382 NSEVL-----EESQSTSLTFG-----398
Db 423 GSDVMRPDYLPSHRHSAVIPPYRPTPDYETVMKQLNRGLVHAERQSHSLRNLNIGSSYA 482
QY 399 -----EGAESPGQSCRRGKEPKVSAAGEPSPAPRSPAGNKQADGA 443
Db 483 YSRPAALVYSQPEIREHAQLPSPAAAHCFSLSYSPHSPSPYPAERRPVGVAVSPEL 542


```
Db 12 SSLKR--KOPKTTFTVRIVTM-----DAEMFNCEMKWKGLDFDLVLCRTLGLRETWFFG 63
QY 83 LEPPDHKKITVWLDLLKPIV-KOIRPKHVVVVKVFFPPD-HTOLOEELTRYLFALOV 140
Db 64 LQY-TIKDTVAVLKMOKKVLHDVSKKEEPTFTHFLAKFYFENAEELVQITQHLFFLQV 122
QY 141 KQDLAQRRLTCNDTSAAALLSHIVSQEIGDFDEAL-DREHLAKNKYIPQO-----D 190
Db 123 KQILDEKIVCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQOELLPKRVINLYQMTPE 182
QY 191 ALEDKIVEFHNIHQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERTAMTAERGRARDEAEYLIKIAQDLEMYGVNYFAIRNKKGTTELLGVDAALGLH 242
QY 251 VFOGFTKIN--AFNNAKVRLKSFKRKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSDEFTIK-----PLDKKIDVFNSSKLRVNLILQ 298
QY 308 ICVEHHAFFRLFEKPKPKPVLFSGSSFRFGRTQKQVLDYVKEGHHKKVQFERKHSK 367
Db 299 LCIGNHDLF--MRRRKADSLEVOOMKAQAREEKARQMERQRLARE---KOMREEAERTR 353
QY 368 IHSIRSLASOPTe--LNSEVLEOSQOSTSLTFEGEASPGQSCRRGCKEKPVSAGBPGSH 425
Db 354 DELERRLLQMKKEATMANEALMRSEETADL-LAEKAI-----TEEEAKLLAQKAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVKVDRTOOSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 9
US-08-628-145-10
; Sequence 10, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; US-08-628-145-10

Query Match 6.1%; Score 335; DB 2; Length 591;
Best Local Similarity 23.8%; Pred. No. 6.3e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERQKPPPTSGKLIVSIKIQMLDDTQFAFEVQAPQAKVLLDVCNHLNLVEGDYFG 82
Db 12 SSLKR--KOPKTTFTVRIVTM-----DAEMFNCEMKWKGLDFDLVLCRTLGLRETWFFG 63
QY 83 LEPPDHKKITVWLDLLKPIV-KOIRPKHVVVVKVFFPPD-HTOLOEELTRYLFALOV 140
Db 64 LQY-TIKDTVAVLKMOKKVLHDVSKKEEPTFTHFLAKFYFENAEELVQITQHLFFLQV 122
QY 141 KQDLAQRRLTCNDTSAAALLSHIVSQEIGDFDEAL-DREHLAKNKYIPQO-----D 190
Db 123 KQILDEKIVCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQOELLPKRVINLYQMTPE 182
QY 191 ALEDKIVEFHNIHQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERTAMTAERGRARDEAEYLIKIAQDLEMYGVNYFAIRNKKGTTELLGVDAALGLH 242
QY 251 VFOGFTKIN--AFNNAKVRLKSFKRKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSDEFTIK-----PLDKKIDVFNSSKLRVNLILQ 298
QY 308 ICVEHHAFFRLFEKPKPKPVLFSGSSFRFGRTQKQVLDYVKEGHHKKVQFERKHSK 367
Db 299 LCIGNHDLF--MRRRKADSLEVOOMKAQAREEKARQMERQRLARE---KOMREEAERTR 353
QY 368 IHSIRSLASOPTe--LNSEVLEOSQOSTSLTFEGEASPGQSCRRGCKEKPVSAGBPGSH 425
Db 354 DELERRLLQMKKEATMANEALMRSEETADL-LAEKAI-----TEEEAKLLAQKAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVKVDRTOOSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 10
US-08-171-718-16
; Sequence 16, Application US/08171718
; Patent No. 5707863
; GENERAL INFORMATION:
; APPLICANT: Trofatter, James A.
; APPLICANT: MacCollin, Mia M.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Tumor Suppressor Gene Marlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,718
; FILING DATE: 22-DEC-1993
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
```

```
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 0609.3850003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-171-718-16

Query Match 6.1%; Score 335; DB 1; Length 595;
Best Local Similarity 23.8%; Pred. No. 6.4e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQKPPPTSGKLVSIKIQMLDDTQAEFVQAPAPGKVLDDAVCNHNLVEGDYFG 82
DB 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKGKDLFDLVCRTLGLRETWFG 63
QY 83 LEFPDHKKITVWDLKPIV-KQIRPKHVVKVVFPPPD-HTQLQBELTRYLFALOV 140
DB 64 LQY-TIKDTVAVLKMCKVLDHVSKEEPTVTFHLAKFYPENAEELVQEIQHLLFLOV 122
QY 141 KQDLAQRGLTCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKVIPOQ-----D 190
DB 123 KQILDDEKICYCPPEASVLLASVAVQAKYGDYPSVHKRGFLAQEELLPRKRVINLYQMTPE 182
QY 191 ALEDKIVEFHNNHIGOTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 183 MWEERTAWYAEHRGRARDEAEWEYLKIAQDLEMVGVNFAIRNKKGTLLGLVDALGLH 242
QY 251 VFQGTFTKIN---AFNWKVKLSFKRRLIKLPDANSAYQDTLEFLMASRDFCKSFVK 307
DB 243 IYDPENRLTPKISFPWNEIRNISYSKTEFTIK----PLDKKIDVFKNSKLRVKNLIQ 298
QY 308 ICVEHHAFFRLFEPEKPKPVLFSGSSFRSGRTQKQVLDYVKGEGHKKVQFERKHSK 367
DB 299 LCIGNHDLF--MRRRKADSLVQOMKAQAREKARKOMERQRLARE---KOMREAEARTR 353
QY 368 IHSIRSLASQPTF--LNSVLEQSQOSTSLTFEGEASPGQSCRRGKPKVSAGEPGSH 425
DB 354 DELERRLLQMKKEATMANEALRSEETADL-LAEKRAQI-----TEEEAKLLAQKAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVEVVVKORTQOSK 463
DB 406 EDEMORIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 11
US-08-478-087-16
; Sequence 16, Application US/08478087
; Patent No. 6077685
; GENERAL INFORMATION:
; APPLICANT: Trofater, James A.
; APPLICANT: MacCollin, Mia M.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
```

```
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,087
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 330
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,718
; FILING DATE: 22-DEC-1993
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 0609.3850003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-087-16

Query Match 6.1%; Score 335; DB 3; Length 595;
Best Local Similarity 23.8%; Pred. No. 6.4e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQKPPPTSGKLVSIKIQMLDDTQAEFVQAPAPGKVLDDAVCNHNLVEGDYFG 82
DB 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKGKDLFDLVCRTLGLRETWFG 63
QY 83 LEFPDHKKITVWDLKPIV-KQIRPKHVVKVVFPPPD-HTQLQBELTRYLFALOV 140
DB 64 LQY-TIKDTVAVLKMCKVLDHVSKEEPTVTFHLAKFYPENAEELVQEIQHLLFLOV 122
QY 141 KQDLAQRGLTCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKVIPOQ-----D 190
DB 123 KQILDDEKICYCPPEASVLLASVAVQAKYGDYPSVHKRGFLAQEELLPRKRVINLYQMTPE 182
QY 191 ALEDKIVEFHNNHIGOTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 183 MWEERTAWYAEHRGRARDEAEWEYLKIAQDLEMVGVNFAIRNKKGTLLGLVDALGLH 242
QY 251 VFQGTFTKIN---AFNWKVKLSFKRRLIKLPDANSAYQDTLEFLMASRDFCKSFVK 307
DB 243 IYDPENRLTPKISFPWNEIRNISYSKTEFTIK----PLDKKIDVFKNSKLRVKNLIQ 298
QY 308 ICVEHHAFFRLFEPEKPKPVLFSGSSFRSGRTQKQVLDYVKGEGHKKVQFERKHSK 367
DB 299 LCIGNHDLF--MRRRKADSLVQOMKAQAREKARKOMERQRLARE---KOMREAEARTR 353
QY 368 IHSIRSLASQPTF--LNSVLEQSQOSTSLTFEGEASPGQSCRRGKPKVSAGEPGSH 425
DB 354 DELERRLLQMKKEATMANEALRSEETADL-LAEKRAQI-----TEEEAKLLAQKAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVEVVVKORTQOSK 463
```

Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 12

US-08-179-738-3
; Sequence 3, Application US/08179738
; Patent No. 5578462
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/179,738
; FILING DATE: 10-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
US-08-179-738-3

Query Match 6.1%; Score 335; DB 1; Length 596;
Best Local Similarity 23.8%; Pred. No. 6.4e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQPPPTSGKLVSIKIQMLDDTQAEFVQAPGKVLDDAVCNHNLNVEGDYFG 82
Db 12 SSLKR--KQPKTFTVRITM-----DAEMEFNCEMKWKGLFDLCVRLGLRETWFFG 63
QY 83 LEFPDHHKTIWLDLLKPIV-KQIRPKHVVKVVFPPD-HTQLOBELTRYLALQV 140
Db 64 LQY-TIKDTVAVLKMDDKVLDDHVDVSKPEPTVTFHLAKFYFENAEELVQEIOTHLFFLQV 122
QY 141 QDLAQRGLTCDNTSAALISHIVQSEIGDFDEAL-DREHLAKNKVIPPQ-----D 190
Db 123 KQILDEKIVCPPEASVLLASVYQAKYGDYDPSVHKRGLAQEELLPKRVINLYQWTE 182
QY 191 ALEDKIVFHHNHIGTPAESDQLLEIARRLMEYGIHLHPAKDRGTKINLAVANTGIL 250
Db 183 MWEERITAWAEHRGRARDEAEYLLKIQADLEMYGVNFYAIRNKKGTTELLGVDAIGLH 242
QY 251 VFQGFTKIN--AFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWK 307
Db 243 IYDPENRLTPKISFPNWEIRNISYSDEFTIK-----PLDKKIDVFKNSSKRVNKLILQ 298

QY 308 ICVEHHAFFRLFEPEPKPKPVLFSSRGSGFRFSGRTQKQVLDYVKEGGHKVKQFERKHSK 367
Db 299 LCIGNHDLF--MRRRKADSLSEVQOMKAQAREEKARQMERQRLARE---KOMREEAERTR 353
QY 368 IHSIRSLASQPTB--LNSEVLEQSQSTSLTFGEAGAESGGQSCRRGKPKVKSAGEPGSH 425
Db 354 DELERLLQMKKEATMANEALMRSEATDL-LAEKQAI-----TEEEAKLLAQAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVEVVKDRTOQSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 13

US-08-628-145-3
; Sequence 3, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
US-08-628-145-3

Query Match 6.1%; Score 335; DB 2; Length 596;
Best Local Similarity 23.8%; Pred. No. 6.4e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQPPPTSGKLVSIKIQMLDDTQAEFVQAPGKVLDDAVCNHNLNVEGDYFG 82
Db 12 SSLKR--KQPKTFTVRITM-----DAEMEFNCEMKWKGLFDLCVRLGLRETWFFG 63
QY 83 LEFPDHHKTIWLDLLKPIV-KQIRPKHVVKVVFPPD-HTQLOBELTRYLALQV 140
Db 64 LQY-TIKDTVAVLKMDDKVLDDHVDVSKPEPTVTFHLAKFYFENAEELVQEIOTHLFFLQV 122

QY	141	KQDLAQRITCNDTSAALLISHIVOSEIGDFEAL--DREHLAKNKYIPQO-----D	190
Db	123	KKQILDEIKVCPPEASVLLASYAAQKYGDYPDVHVKRGFLAQEELLPKRVINLYQMTP	182
QY	191	ALEDKIVEPHHHIGOTPAESDFOLLEIARRLEMVGIRLHPAKDRGCKINLAVANTGIL	250
Db	183	MWEERTATWAEHRGARDEAEMEXYLIAQDLEMVGUNYFAIRNKKGTETLLGVDAALGIH	242
QY	251	VFOGFWKIN--AFNNAWKRLSPFKRKLILRPDANSAYODTLFELMASDFOCPSWK	307
Db	243	IYDPENRLTPKISFPNEIRNISYSKEETIK----PLDKKIDVTKFNSSKLRUVKNLILO	298
QY	308	ICVEHHAFRLFEPPKPVPYLSRGSFSFGRTQKVLDYVKGEGHKVKQFERKHKS	367
Db	299	LCIGNHDLP--MRRRKADSLLEVQQMQKAQAREKARKQMERQLARE---KQMRBAERTR	353
QY	368	IHSIRLASOPTe--LNSEVLQSOSTSLTFEGEAESPGGOSCRGKPEKPYASAGEPSSH	425
Db	354	DELERRLLQMKEEATMANEARMBSETADL-LAEKRAQI-----TEEEAKLLQAQAAEA	405
QY	426	PSPAPRRSPAGNKQADGAASAPTEEBSVVVKDRTQOSK	463
Db	406	EOMORIK-----ATAIRTEEKRLMEOKVLEAE	434

```

RESULT 14
US-08-179-738-7
; Sequence 7, Application US/08179738
; Patent No. 5578462
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; City: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/179,738
; FILING DATE: 10-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
; US-08-179-738-7

```

Query Match 6.1%; Score 332; DB 1; Length 584;
Best Local Similarity 23.6%; Pred. No. 1.1e-19;
Matches 108; Conservative 108; Mismatches 190; Indels 52; Gaps 15;

```

23  STLERGQKPPPTSGKLVSIKIOMLDDTQAEVFPQAPBGKVLDDAVCNHNLNVEGDFVG 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
12  SSLK--KQPKTFTVITM-----DAEMFNCENMKWKGDFDLVCTRLGLRDTWFFG 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
83  LEPPDHKKITVWLDDLLKPTV-KOIRPKHVVKVVKFFPPD-HTQLOBELTRYLFALQV 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64  LQ-TTKDVTAVMLKMDKKDHDHDSKEEPTTHFLAKFYPENAEELVQEIITQHLFFLQV 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
141  KODLAGRLTCNDTSAALLISHIVSEIGDFDEAL-DREHLAKNKYIQQ-----D 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
123  KKEILDEKVCYCPPEASVLLASVAKYGDYDPSVHKRGFLAQEELLPKRVINLYQMTPE 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
191  ALEDKTVEFHNNHIGOTPAESDFOLLEIARRLEMGIURLHPAKDREGTKINLAVANTGIL 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183  MWEERTITAYAEHGRGARDEAEMEXLIKIAQDLEMVGVNYFTIRNKKGTLELGVDAALGLH 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251  VFQGFYKIN---AFNMAKYKLSFKRKRFLIKLRPDANSAYOOTLEFLMASRDFCKCFWK 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243  IYDPENRLTPKISFPWNEIRNYSDKETIK---PLDKKIDVPKFDSSKLRVKNLLIQ 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
308  ICVEHHAFTLFEPPKPKPKPVLSRGSSFRFSGRTOQVLDYVYKGGHKKVQFPERKHSK 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
299  LCIGNHDLF--NRRRKADSLSEVQOMKAQAREEKARKQMERQRLARE---KOMREEAERTR 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
368  IHSIRLSAQPT--LNSEVLEQSOOSTSLTTEGEAESPGGOSCRGRKPKYSAGEPGSH 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
354  DELERLLQMKBEATMANALMRSEETADL-LAEKAQI-----TEEEAKLULAQAAEA 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
426  PSPAPRRSPAGNKQADGAASAPTEEEEEEVVKDRTQOSK 463
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
406  EQEMQRIK-----ATAIRTEEEKRLMEQKVLEAE 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-08-628-145-7
; Sequence 7, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids

```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
;
US-08-628-145-7

Query Match          6.1%; Score 332; DB 2; Length 584;
Best Local Similarity 23.6%; Pred. No. 1.1e-19;
Matches 108; Conservative 108; Mismatches 190; Indels 52; Gaps 15;

QY 23 STLERQKPPPTPSGKLVSIQMLDDTQAEFEVQPARAFKVLDDAVCNHLNLVEGDYFG 82
Db 12 SSLKR--KQPKFTVRIVTM-----DAEMEFNCENKWKGDLDLVCRITGLRDTWFFG 63

QY 83 LEPFDHKKITVLDLLKPIV-KQIRPKHVVKVFFVKKFPPD-HTQLQEEELTRYLFALQV 140
Db 64 LQY-TIKDTVAMLKMDKKVLDHDKVEEPTVTHFLAKFYPENAEELVQEIQTQHLFFLQV 122

QY 141 KDLAQGRLTCDNTSAALLSHIVQSEIGDFEAL-DREHLAKNKVIPOQ-----D 190
Db 123 KKEILDEKVCYCPPEASVLLASVAVQAKYGYDPSVHKRGFLAQEELPRVINLYQMTPE 182

QY 191 ALEDKIVEFHNNHIGTPAESDFQLLEIARRLEMVYGIHLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERITAWAEHRGRARDEAEWYELKIAQDLEMVGVNYFTIRNKKGTETLLGLVDALGLH 242

QY 251 VFOGFTKIN---AFNWKVRKLSFKKRFLIKLRPDANSAYQDTLEFLMASRDFCKSWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSDKFTIK----PLDKKIDVFKFDSKLRVNNKLIQ 298

QY 308 ICVEHHAFFRLPEEPKPKPVLFSGSSFRSGRTQKOVLDVYKGGHKKVQFERKHSK 367
Db 299 LCIGNHDLF--MRRRKADSLVQOQRAQAREKARKOMERQRIARE---KOMREAEERTR 353

QY 368 IHSIRSLASOPTF--LNSEVLEQSQOSTSLTFEGEAGSPGQSCRRGKPKVKSAGEPGSH 425
Db 354 DELERRLLQMKKEATWANEALMRSEETADL-LAEKQAQI-----TEEEAKLLAQKAAEA 405

QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVEVVKDRTQOSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434
```

Search completed: December 6, 2001, 08:55:08
Job time: 290 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:46:28 ; Search time 44.17 Seconds
(without alignments)
1802.183 Million cell updates/sec

Title: US-09-555-342A-2
Perfect score: 5463
Sequence: 1 MGEIEQRPTGSLGAPENS.....SATSSASRPHVLHSHKESLVY 1045

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5456	99.9	1045	JC5795	CDEP protein - hum
2	850.5	15.6	1111	T23047	hypothetical prote
3	709.5	13.0	850	MMHUE4	erythrocyte membra
4	698	12.8	926	A41105	protein-tyrosine-p
5	692	12.7	858	A45613	protein 4.1, P4.1
6	691.5	12.7	1698	T13800	coracine gene prote
7	686.5	12.6	230	T34541	hypothetical prote
8	681	12.5	801	A37353	membrane protein 4
9	644	11.8	1026	T19631	hypothetical prote
10	568	10.4	568	T25859	hypothetical prote
11	541	9.9	913	A41109	protein-tyrosine-p
12	475.5	8.7	554	J00188	membrane protein 4
13	465.5	8.5	414	T22498	hypothetical prote
14	457.5	8.4	4549	T20771	hypothetical prote
15	457.5	8.4	4667	T20774	hypothetical prote
16	441.5	8.1	961	A25380	faciogenital dyspl
17	434	7.9	1175	S31005	protein-tyrosine-p
18	430	7.9	639	T20772	hypothetical prote
19	414.5	7.6	875	T19678	hypothetical prote
20	401.5	7.3	1176	T196345	protein tyrosine p
21	396	7.2	138140	I38140	protein-tyrosine-p
22	391.5	7.2	577	S39804	moesin - pig
23	390.5	7.1	577	A41289	moesin - human
24	386.5	7.1	583	A41129	radixin - mouse
25	385.5	7.1	583	S39805	radixin - pig
26	382.5	7.0	583	A46127	radixin - human
27	367.5	6.7	630	T47177	hypothetical prote
28	365	6.7	581	T45889	ezzrin - bovine
29	356.5	6.5	586	A34400	ezzrin [validated]

30 356 6.5 586 1 B41129 ezzrin - mouse
31 349 6.4 654 2 T30957 hypothetical prote
32 341.5 6.3 1187 1 JC4155 protein-tyrosine-p
33 337 6.2 563 2 T29262 hypothetical prote
34 335 6.1 595 2 S33809 neurofibromin 2 -
35 334 6.1 591 2 I54368 merlin protein - m
36 334 6.1 596 2 I68664 merlin - mouse
37 333.5 6.1 1189 1 JC2366 protein-tyrosine-p
38 331 6.1 564 2 T29263 hypothetical prote
39 315 5.8 2490 1 A54971 protein-tyrosine-p
40 300 5.5 559 2 A45620 cyto villin homolog
41 299 5.5 559 2 S49143 EG10 protein - tap
42 296.5 5.4 2466 2 I67629 protein tyrosine p
43 294.5 5.4 2294 2 I67630 protein tyrosine p
44 293 5.4 2450 2 S71625 protein-tyrosine-p
45 291 5.3 1737 2 A59235 unconventional myo

ALIGNMENTS

RESULT 1
JC5795
CDEP protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C:Accession: JC5795
R:Koyano, Y.; Kawamoto, T.; Shen, W.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.
Biochem. Biophys. Res. Commun. 241, 369-375, 1997
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain
ng factors.
A:Reference number: JC5795; MUID:98086358
A:Accession: JC5795
A:Molecule type: mRNA
A:Residues: 1-1045 <KOY>
A:Cross-references: DBJ:AB008430
C:Comment: This protein is involved in the adhesion, proliferation, and differentiation
C:Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain homolo
F:1-374/Domain: ezzrin-like #status predicted <E2R>
F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>
F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 99.9%; Score 5456; DB 2; Length 1045;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGEIEQRPTGSLGAPENSIGSTLERGQKPPPTPSGKLVSIKIQMLDDTQEAFAEVPQRA 60
DB 1 MGEIEQRPTGSLGAPENSIGSTLERGQKPPPTPSGKLVSIKIQMLDDTQEAFAEVPQRA 60
QY 61 PGKVLDDAVCNHNLNVEGDFYFGLFPPDHKKITVWLDLLKPIVKQIRRPKHVVVKKVVF 120
DB 61 PGKVLDDAVCNHNLNVEGDFYFGLFPPDHKKITVWLDLLKPIVKQIRRPKHVVVKKVVF 120
QY 121 PPDHTQLQELTRYLFALQKQDLAQGRLLTNDTSAALLISHIVQSEIGDFDEALDREHL 180
DB 121 PPDHTQLQELTRYLFALQKQDLAQGRLLTNDTSAALLISHIVQSEIGDFDEALDREHL 180
QY 181 ANKNTYPOQDQALDKIVFHHNHIGQTPAESDFQLLEIARRLEMYGIRLHPAKDREGTKI 240
DB 181 ANKNTYPOQDQALDKIVFHHNHIGQTPAESDFQLLEIARRLEMYGIRLHPAKDREGTKI 240
QY 241 NLAVANTGILVFGQFTKINAFNNAWKVRKLSFKRRKRLIKLRPDANSAYODTLEFLMASRD 300
DB 241 NLAVANTGILVFGQFTKINAFNNAWKVRKLSFKRRKRLIKLRPDANSAYODTLEFLMASRD 300
QY 301 FCKSFWKICVEHHAFFRLFEPPKPKPVLFSGSSFRFSGRTQKQVLDYVREGGHHKYQ 360
DB 301 FCKSFWKICVEHHAFFRLFEPPKPKPVLFSGSSFRFSGRTQKQVLDYVREGGHHKYQ 360
QY 361 FPKKSHKIHISRLASQPTLELSEVLEQSQOSTSLTFGGAESPGQSGCSRKPEKPVKAG 420
DB 361 FPKKSHKIHISRLASQPTLELSEVLEQSQOSTSLTFGGAESPGQSGCSRKPEKPVKAG 420

Db 361 FERKSHKSIHSIRASQPTLNSEVLEQSQOOSTSLTFGEAESPGQSCRRGKEPKVSAG 420
 QY 421 EPSCSPSPARRSPAGNKAQDGAASAPTEEEVWVWDRTOQSKPQPQSTGLTGSPLH 480
 Db 421 EPSCSPSPARRSPAGNKAQDGAASAPTEEEVWVWDRTOQSKPQPQSTGLTGSPLH 480
 QY 481 SELSVNSQSGVAPANTYLSPLNPSDPKQASPLISPLLNQACPRTDDEDEGRKRPTDK 540
 Db 481 SELSVNSQSGVAPANTYLSPLNPSDPKQASPLISPLLNQACPRTDDEDEGRKRPTDK 540
 QY 541 AYPFAKEVSTTERTYKLELVITSNFQSTVSKEDAMPEALKSLIFNPFPLHKHFTNFKL 600
 Db 541 AYPFAKEVSTTERTYKLELVITSNFQSTVSKEDAMPEALKSLIFNPFPLHKHFTNFKL 600
 QY 601 EIQORLALWEGRSNAOIRYQRTGDVYMLKNIQGMKHLAHLAKHSEALENGIKSSRR 660
 Db 601 EIQORLALWEGRSNAOIRYQRTGDVYMLKNIQGMKHLAHLAKHSEALENGIKSSRR 660
 QY 661 LENFCRDFELQKVCYLPLNTFLLRPLRLMHYKQVLERLCKKHPPSHADFRDCRAALAEI 720
 Db 661 LENFCRDFELQKVCYLPLNTFLLRPLRLMHYKQVLERLCKKHPPSHADFRDCRAALAEI 720
 QY 721 TENVAOLHGTMIKMFQKLELKDOLIGDNLVPGREFIRLGSLSKLSGKLOQRMFF 780
 Db 721 TENVAOLHGTMIKMFQKLELKDOLIGDNLVPGREFIRLGSLSKLSGKLOQRMFF 780
 QY 781 LFNDVLLYTSRGLTASNOFKVHGPOLYGMTIEESEDWGVPHCLTLRGOROSIIVAAS 840
 Db 781 LFNDVLLYTSRGLTASNOFKVHGPOLYGMTIEESEDWGVPHCLTLRGOROSIIVAAS 840
 QY 841 RSEMERKVEDIQMAIDLAESKSPAPFEFLASSPPDNKSPDEATAADOSEDDLSASRTSL 900
 Db 841 RSEMERKVEDIQMAIDLAESKSPAPFEFLASSPPDNKSPDEATAADOSEDDLSASRTSL 900
 QY 901 EROAPHRGNTMWHVCHWRNTSVSWDFSTAVENQLSGNLLRKFKNNGWOKLWVFTNFC 960
 Db 901 EROAPHRGNTMWHVCHWRNTSVSWDFSTAVENQLSGNLLRKFKNNGWOKLWVFTNFC 960
 QY 961 LFPYKSHQDNHPLASPLPLGYSLTIPSESENQKDVVKLHFKSHVYYPFAESEYTFERW 1020
 Db 961 LFPYKSHQDNHPLASPLPLGYSLTIPSESENQKDVVKLHFKSHVYYPFAESEYTFERW 1020
 QY 1021 MEVIRSATSSASRPHVLSHKSLEY 1045
 Db 1021 MEVIRSATSSASRPHVLSHKSLEY 1045
 RESULT 2
 T23047
 hypothetical protein H05G16.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T23047
 R:White, S.
 submitted to the EMBL Data Library, June 1997
 A:Reference number: Z19661
 A:Accession: T23047
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1111 <WIL>
 A:Cross-references: EMBL:Z97190; PIDN:CAB10024.1; GSPDB:GN000028; CESP:H05G16.1
 A:Experimental source: clone H05G16
 C:Genetics:
 A:Gene: CESP:H05G16.1
 A:Map position: X
 A:Introns: 97/3; 139/2; 171/3; 213/2; 294/3; 348/2; 373/2
 Query Match 15.6%; Score 850.5; DB 2; Length 1111;
 Best Local Similarity 26.6%; Pred. No. 1.3e-41;
 Matches 287; Conservative 157; Mismatches 363; Indels 271; Gaps 38;
 QY 22 ISTLERGQRPPTPS-----GKLVSIKIQMLDDTQEAPE----- 55

Db 1 MSNIPRGVAGAPPPMNSKRGRLVCIKVRMLDDTVAVFHLGVICFFCFVVSFHTYVEL 60
 QY 56 VQORAPKVLDAVCHNLNVECDYFGLFPHDKKITVWLDLLKPIVKOIRR- PKHVVK 114
 Db 61 LQHKAIQTLLDEVCHNLNVECDYFGLFIDINGNHCWLDREKTLTLRTNGSTDAKPY 120
 QY 115 FVVKFPDPDHTLOELTRFLFALOVKQDLAOGRLTCNDTSAALLISHIVQSEIGDF--D 172
 Db 121 FVVKYTPNPIDLEEYTRYLFTMQIKRDLALGELHCSNTASLLSAYLVQSECGDFSSE 180
 QY 173 EALDREHLAKNKYIPQD--ALEDKIVFHHHIGOTPAESDFOLLEIARLEMYGIRLHP 231
 Db 181 DYPDATYLSHTFRVNPQTLFQKVMNDNRNFTGMTGESDLAMLEAVARRCDFYGVKHLA 240
 QY 232 AKDREGTKINAVANTGILVFOGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDT 291
 Db 241 AKDIDGNDAAALSVMLGIKIVFQLOLDITTFWARIKLSFKRKKLLVLRHPSYQYLKET 300
 QY 292 LEFLMASRDECKSFVKICVEHHAFFRLPEPKPKPVLF--SRGSSFRSGRTQKQVLDY 350
 Db 301 VEFSETRDECKNFVKCVEHHAFFRCVQAEPEKRETFISGSSFRHGTQKQVLDY 360
 QY 351 VKEGGHKKVQFER----- 368
 Db 361 VREHKKRREPFTPLRSAASTRKGTYSSTYGLVTDRTKRNKSVYEPNQTDPYNKHQWT 420
 QY 369 HS-----IRSLASQPT-----LNSEVLEQ--SQOSTSLTFGEAESPGQSCRRG-- 412
 Db 421 HSMPIAHIISSQADHSGTLDARVGSVTRDPTSVTQHLRQLKRSERCISSVDIV 480
 QY 413 KSPKVSAGPGHSPAPRRSPAGNKAQDGAASAP--EVEEVVWDRTOQSKPQPQPS 470
 Db 481 ERQSRSCRPVSHAINI--SSTSSKEQPDLSVPLNVLSDDLQMVCKEIEIQNDPKSV 538
 QY 471 TG-----SLTGSPLSELVNSOGGVAP--ANVTLSPLNLSPTDKQAS 510
 Db 539 SGDNFQRRSSRDYDNVEDSYRLSDH--ERSTRSEVGVGSKFAAATIFNSTFVARPKGS 596
 QY 511 PLISPLL--NDQACPRDDEDEGRK-----RPTDKAYFAKE----- 547
 Db 597 NVVKRVVAHTKSTPNTDDEEGALKSASQYQFRHIKEYPFARNANIVIEIDGPNVDLS 656
 QY 548 -----VSTERTYKLELVITSNFQSTVSKEDAMPEALKSLIFN-PEPLHKHFTNLF 599
 Db 657 ARRSAPATTTTTRTYANTTTTTSKVLTSAGVL--MKPKVISNDHEPSSGTHSS-- 711
 QY 600 KEIQRALWEGRSNAOIRYQRTGDV-----MLKNIQGMKHLAHL----- 640
 Db 712 -----RVSPESGTYGALGPLGRVITKENMMITPEGFKEKKAKPNPPKPSPP 759
 QY 641 -LWKHSEALEALENGIKSSRRLENFCRDFELQKVCYLPLNTFLLRPLRLMHYKQVLERL 699
 Db 760 VOAVHSETEVRE--IKTERASS-----PDVQKCHLFNSD-----IP-- 791
 QY 700 CKHPPSHADFRDCRAALAEITEMVAQLHGTMIKMFQKLEHKLKDLIGDNLVVPGR 759
 Db 792 FSRAPLKPEDKKERKSSLSRPAISVQSEDN---PDVQKCHLFNSD-----IP-- 836
 QY 760 FTRLSKLSKLGKLOQRMFFLNDVLLYTSRG-----LTASQFKVHQLPL---- 807
 Db 837 ----YTLTMRNVENTQSLPYSSFKDVKSTAKQGYESNSLKRYSKSPFRRKRSLOLVPRK 892
 QY 808 -----YGMTIEESEDWGVPHCL-----TLRGOROSIIVAASRSEMEKKEWEDIQM 853
 Db 893 RLPSGNTSAQDHTISPTTPDSVLELRLRRSLLGDK-SVTKTKSRDTPRQTPVR- 950
 QY 854 AIDLAESKSPAPFEFLASSPP-----DNKSPDEATAADOSEDDLSASRTSLERQAPH 906
 Db 951 -FDLPPSPCSP--AGGSTPFIISILNDOLFECVSESRSLEDMD-----RLDKTTTPH 1000
 RESULT 3

RESULT 7

T34541

hypothetical protein DKFZp434M2221.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, October 1999

A:Reference number: Z21541

A:Accession: T34541

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-230 <BLU>

A:Cross-references: EMBL:AL122052

A:Experimental source: adult testis; clone DKFZp434M2221

C:Genetics:

A:Note: DKFZp434M2221.1

Query Match 12.6%; Score 686.5; DB 2; Length 230;

Best Local Similarity 63.8%; Pred. No. 4.7e-33;

Matches 132; Conservative 29; Mismatches 39; Indels 7; Gaps 3;

QY 831 RQSTIVAAASRSSEMEKVVEDIQAIDLAEKSSSPAPEF-----LASSPPDNKNSPDEATAAD 886

DB 3 QKTVIAAASRLERKEKWMKLDLNSAIQAQKSGGDTAPALPGRTVCTTRPP--RSPNE-VSL 59

QY 887 QESEDLLASRTSLERQAPHRGNTMVHVCVHRNTSVSMVDFSIAVENQLSGNLLRKFKNS 946

DB 60 QESEDDARGVRSSLEGHGCHQRANTTMVCHVYRTSVSRADHSAAVENQLSGYLLRKFKNS 119

QY 947 NGWQKLVWVFNTFCFLFYKSHQDNHPLASLPLLGYSITPSESENTQKYVFKLHKFSHV 1006

DB 120 HGWQKLVWVFNTFCFLFYKTHQDDYPLASLPLLGYSVSIPIREADGIHKDYVFKLQFKSHV 179

QY 1007 YFRAESEYTFERWMEVIRSATSSASR 1033

DB 180 YFRAESKYTFERWMEVIOGASSAGR 206

RESULT 8

A37353

membrane protein 4.1 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 13-Aug-1999

R:Spencer, M.; Giebelhaus, D.H.; Kelly, G.M.; Bicknell, J.; Florio, S.K.; Milam, A.H.

Dev. Biol. 139, 279-291, 1990

A:Title: Membrane skeleton protein 4.1 in developing Xenopus: expression in postmitot

A:Reference number: A37353; MUID:90249600

A:Accession: A37353

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-801 <SPE>

A:Cross-references: GB:M20621; NID:g214090; PIDN:AAA49695.1; PID:g214091

R:Giebelhaus, D.H.; Eib, D.W.; Moon, R.T.

Cell 53, 601-615, 1988

A:Title: Antisense RNA inhibits expression of membrane skeleton protein 4.1 during em

A:Reference number: A29901; MUID:88223353

A:Accession: A29901

A:Molecule type: mRNA

A:Residues: 1-550 <GIE>

C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology

C:Keywords: cytoskeleton; membrane protein

F:195-470/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 12.5%; Score 681; DB 2; Length 801;

Best Local Similarity 30.7%; Pred. No. 6.1e-32;

Matches 180; Conservative 80; Mismatches 212; Indels 114; Gaps 15;

C:Species: Mus musculus (house mouse)
C:Date: 03-Feb-1994 #sequence_revision 20-Aug-1994 #text_change 05-Nov-1999
C:Accession: I55505; J00188
R:Takeuchi, K.; Kawashima, A.; Nagafuchi, A.; Tsukita, S.
J. Cell Sci. 107, 1921-1928, 1994
A:Title: Structural diversity of band4.1 superfamily members.
A:Reference number: I55505; MUID:95074267
A:Accession: I55505
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-554 <RES>
A:Cross-references: GB:D28818; NID:9465547; PIDN:BAA05978.1; PID:9465548
C:Superfamily: protein 4.1 membrane-binding domain homology
F:13-295/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 8.7%; Score 475.5; DB 2; Length 554;
Best Local Similarity 30.3%; Pred. No. 3.3e-20;
Matches 135; Conservative 79; Mismatches 182; Indels 49; Gaps 14;

QY 43 KIOMLDDTQAEFVQK-----APGKVLDAVCNHLNVEGDYFGLFEPDPKKTIVWLDL 97
DB 14 EVLLDESKTLTQOQGIKSKTSGVLDHVFRRHNLVEIDFGLRYCDRSHQTVLDP 73
QY 98 LKPIVQK---IRRPKHVVVVFVFPDPDHTQLOELTRYLFALQVQDLAQGLTCNDT 154
DB 74 AKTLAEHKLINTGPPVTLVFGIKFYAEDPCCKLEETRYQFFLVQVQDALQGLRCPVN 133
QY 155 SAALLSHIVQSIGDFDEALDRE-HLAKNKYIP-QQDALEDKIVFEHNNHICQTAE 212
DB 134 IAAOMGAYATQAEQDHPDKHTAGVSEYFVDPDQKEELEAIEIRHKTLMQAPSEAE 193
QY 213 FQLELTARRLEMGIRLHPAKDREGTKINLAVANTGLVFOGTFKINAFNAKVRKLSPK 272
DB 194 LNLRTAKSLEMGVLDHPVYGVNGENKSEYFGLTPSGVVVYTKNKKQVGFWPRTKTVHK 253
QY 273 KRRLIK-LRPDANSAYQDTLEFLMASRDFCSFKVFCVHEHAFRRLEFEPKPKPVLV 331
DB 254 ETQFLRVLGKDCN---ETSFEEARSKTACKHLWKCSVEHHTFFRMPDTESNLSRKLS 310
QY 332 SRGS-SFRFSGRQKV-LDYVEGHHKKVQFERKISKTHSIRSLASQPTLSEVLEQS 389
DB 311 KFGSISYKHYRTALQMSRDLSQLPRNQNVVRSKTYPKRVAQTPTGSNN--INRI 368
QY 390 QQSTSLTFEGG-----AESPGGSCRGKPKVYSGEPGSHSPAPRRSPAGNKQADGAA 444
DB 369 TANTENGEGTKITLAPSP-VKSPKAK-----NENSPDPQRS-----KS 408
QY 445 SAPTEEE-----EYVVKDRTOQSK 463
DB 409 HAPWEENGPOSGLYNSSSDRTKSPK 433

RESULT 13
T22498
hypothetical protein F52D10.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T22498
R:Kershaw, J.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19571
A:Accession: T22498
A:Molecule type: DNA
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-414 <WIL>
A:Cross-references: EMBL:Z66564; PIDN:CAA91472.1; GSPDB:GN000028; CESP:F52D10.6
A:Experimental source: clone F52D10
C:Genetics:
A:Gene: CESP:F52D10.6
A:Map position: X
A:Introns: 46/2; 195/3; 264/1; 337/3; 389/2

Query Match 8.5%; Score 465.5; DB 2; Length 414;
Best Local Similarity 28.0%; Pred. No. 8.2e-20;
Matches 119; Conservative 87; Mismatches 176; Indels 43; Gaps 10;
QY 623 IGDVMLKNTQGMKHLAAHLWKHSEA-LEALENGIKSSRRLENFCRDFELQKVCYLPLNTF 681
DB 2 VAPVLLATVHNLTPLYQSLVQNFPIYVSALDQLYRSKIPFRTIMSKFESKECYTQVNW 61
QY 682 LLRPLRLMHYKOVLERLCKHHPPSHADPDC-----RAALAEITEMVAQLHGTMI 732
DB 62 LKILNRLINWQVPLARVIE-----IQSECGNDTETAFGVAMDKILIEFAKTKATRQ 115
QY 733 KMFNFOLKHLKDK--LIGIDNLVVPGRFIRGLSGLSKLGKLOQRMFFLDFNLLYTS 790
DB 116 SLEEYIHVLQVERDTGLVGI--LTHPNRKILRVGFLRSARRAPCCCRIMVLCSDRILFGH 173
QY 791 RGLTA-SNPFKVGQPLPLVGMTIESEDEW---GVPHCLTLRQROSIIVAASRSEMEK 846
DB 174 RGVNLDGNFTVHAEFKLGMMIDEGDTYKVMGDNENNVITLHNADISIVFAAPDRA---S 230
QY 847 WVEDIOMADLAEKSSPAPFLASSPPDNKSPDEATAADOESEDLLSASRTSLERQAPH 906
DB 231 WIEDITEAKNAARAKIDPLSV-----MEKKEENDIDMTKVLPLESPV 275
QY 907 RGMTVHVGVHRTSVSMVDFSIQSLNQLSGNLLRKFKNKGWKLWVVFNEFLFFYKS 966
DB 276 SKMSPLQICWYKCSFGRKDVNKMITWCGYLKRLNNSGQDLWVVMCCHTLYFYN 335
QY 967 HODNHPASLPLGLYSLTIPSESENIOK-DYVFKLHKSHVYFRAESEYTFERMEVIR 1025
DB 336 HNEREPALHLSMDYGVGLPTVADKIDNHENCFKLYGSHTYFRTDSYFFERWVDSIF 395
QY 1026 SATSS 1030
DB 396 QAAIS 400

RESULT 14
T20771
hypothetical protein ZK270.2a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20771; T27814
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19322
A:Accession: T20771
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4549 <WIL>
A:Cross-references: EMBL:Z81499; PIDN:CAB04090.1; GSPDB:GN000019; CESP:ZK270.2a
A:Experimental source: clone F11C3
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20423
A:Accession: T27814
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4549 <WIL>
A:Cross-references: EMBL:Z82089; PIDN:CAB05003.1; GSPDB:GN000019; CESP:ZK270.2a
A:Experimental source: clone ZK270
C:Genetics:
A:Gene: CESP:ZK270.2a
A:Map position: 1
A:Introns: 61/3; 112/1; 244/3; 328/2; 384/1; 406/3; 4302/3; 4343/3; 4377/2; 4469/3; 4

Query Match 8.4%; Score 457.5; DB 2; Length 4549;
Best Local Similarity 28.1%; Pred. No. 8e-18;
Matches 158; Conservative 83; Mismatches 239; Indels 83; Gaps 20;
QY 28 GOKPPPTPS-----GKLVSIIQMLDDTQAEFVQRAFGKVLDDAVCNHLNVEGDYFG 82

Db	27	GDQKHPTTSQOPRDSKQMLAKVLLPDGVQVQKEFEVKNKSEGEALFRQVTRDLSIEEREYFS	86
Qy	83	LEFPDHHKIT-VWLDLKLPIVKQIR-RPKHVVVVFVFFPPDHTQLQOEELTYLFAIQV	140
Db	87	LCFYDKDEGRHWLYNDKNIQAQIKGLPWE--FSFEVKEVPTPTTIVDDHARYVFLQL	144
Qy	141	KODLAQGLTACNDPSSAALLISHIVSEIGDFD-EALD--REHLAKNKYIIFOQDALED---	194
Db	145	RRDLTLGLKPATADTHSLLSGFAQIEFGDAPAEWTDAYEQFIVASKLVPSAQNPTYK	204
Qy	195	KIVFHHNHGQTPAESDFOLLEIARLEMYGIRLHPA-KDREGTKINLAVANTGILVFQ	253
Db	205	KIVDLHREMGRQTPSEAEQNFLDHCKHLALYGIHLFRAISDKDKKPPVDVGIGAAGINIYO	264
Qy	254	GFTKINAFNAKVRKLFSEKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFVKICVEHH	313
Db	265	DEQTHSFQWNIKIGYRTYFSIKLACGTVEKNEKTYFLKLPNHVAAKRTWKCAVEHH	324
Qy	314	AFRFLFEEPPKPKPVLFSGSS--FRESGRTQKVLDYVKEGGHKV--QFERKHSKHS	370
Db	325	TFEFLI-QPEDKTHKSFNFGSQRFYQGRTOFT-----KIASQWEDKPSIVDR	373
Qy	371	IRSLASQTELSNVLQSQOSTSLTGE-----GAESPQGQSCRRGKEPKVSAG	420
Db	374	APSAMSQPI---ATAENKQLQTLNLTSELEQROFERYRRALTPKSYTSSRQDDAVSTA	430
Qy	421	-----EPGSHSPAPRR---SPAGNKQADGAASAPTEEEEEEVVKDRTOQSK	463
Db	431	TFAKYSRPSTLLVYSTSGHHQSPHELSYSPRGDSSNYSSAAYYMSERSSL--RTPSSA	488
Qy	464	POPPQST-----GSLTGSPLHSLSVNSGGVAPANVTLSPNL-----SPD	505
Db	489	YYPSEVTSPTSPYVFEGQHGXHVQMORTASTSATGSPNSGARRNLFRASDARSSQA	548
Qy	506	TKQASPLIS---PLINDQACPT	525
Db	549	SRDSVRLVSFHEPVDNPEIPNT	571
RESULT 15			
T20774			
hypothetical protein zk270.2d - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999			
C:Accession: T20774; T27818			
R:Kershaw, J.			
submitted to the EMBL Data Library, November 1996			
A:Reference number: Z19322			
A:Accession: T20774			
A:Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-4667 <ML>			
A:Cross-references: EMBL:Z81499; PIDN:CAB54224.1; GSPDB:GN000019; CESP:ZK270.2d			
A:Experimental source: clone F11C3			
R:Kershaw, J.			
submitted to the EMBL Data Library, November 1996			
A:Reference number: Z20423			
A:Accession: T27818			
A:Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-4667 <MI2>			
A:Cross-references: EMBL:Z82089; PIDN:CAB54513.1; GSPDB:GN000019; CESP:ZK270.2d			
A:Experimental source: clone ZK270			
C:Genetics:			
A:Gene: CESP:ZK270.2d			
A:Map position: 1			
A:Introns: 61/3; 112/1; 244/3; 328/2; 384/1.406/3; 4302/3; 4343/3; 4384/1; 4422/3; 4462/3; 4463/3			

Query Match 8.4%; Score 457.5; DB 2; Length 4667;
Best Local Similarity 28.1%; Pred. No. 8.3e-18;
Matches 158; Conservative 83; Mismatches 239; Indels 83; Gaps 20;

Query Match	8.48:	Score 457.5:	DB 2:	Length 4667:
-------------	-------	--------------	-------	--------------

Best Local Similarity 28.1%; Pred. No. 8.3e-18;
Matches 158; Conservative 83; Mismatches 239; Indels 83; Gaps 20;

Search completed: December 6, 2001, 08:52:34
Job time: 366 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:46:48 ; Search time 28.69 Seconds
(without alignments)
1335.474 Million cell updates/sec

Title: US-09-555-342A-2

Perfect score: 5463

Sequence: 1 MGIEQRPTPGSRIGAPENS.....SATSSASRPVLSHKESLVY 1045

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708.5	13.0	864	1 4L_HUMAN	P11171 homo sapien
2	698	12.8	926	1 PTH4_HUMAN	P29074 homo sapien
3	694	12.7	881	1 E4L1_HUMAN	Q9H490 homo sapien
4	692	12.7	858	1 4L_MOUSE	P48193 mus musculus
5	681	12.5	801	1 4L_XENLA	P11434 xenopus lae
6	644	11.8	1026	1 PTP1_CAEEL	P28191 caenorhabdi
7	626	11.5	732	1 YF48_HUMAN	Q9Hcm4 homo sapien
8	541	9.9	913	1 PTH3_HUMAN	P26045 homo sapien
9	532	9.7	619	1 NBL4_BRARE	O57457 brachydanto
10	518.5	9.5	598	1 NBL4_HUMAN	Q9Hcs5 homo sapien
11	475.5	8.7	554	1 NBL4_MOUSE	P52963 mus musculus
12	450.5	8.2	960	1 FGD1_MOUSE	P52734 mus musculus
13	441.5	8.1	961	1 FGD1_HUMAN	P98174 homo sapien
14	434	7.9	1175	1 PTHL_RAT	Q62728 rattus norv
15	401.5	7.3	1176	1 PTHL_MOUSE	Q62136 mus musculus
16	396	7.2	1174	1 PTHL_HUMAN	Q16825 homo sapien
17	391.5	7.2	576	1 MOES_MOUSE	P26041 mus musculus
18	391.5	7.2	576	1 MOES_PIG	P26042 sus scrofa
19	390.5	7.1	576	1 MOES_HUMAN	P26038 homo sapien
20	386.5	7.1	583	1 RAD1_MOUSE	P26043 mus musculus
21	385.5	7.1	583	1 RAD1_PIG	P26044 sus scrofa
22	382.5	7.0	583	1 RAD1_HUMAN	P35241 homo sapien
23	365	6.7	580	1 EZR1_BOVIN	P31976 bos taurus
24	356.5	6.5	585	1 EZR1_HUMAN	P15311 homo sapien
25	356	6.5	585	1 EZR1_MOUSE	P26040 mus musculus
26	341.5	6.3	1187	1 PTNE_HUMAN	Q15678 homo sapien
27	335	6.1	595	1 MERL_HUMAN	P35240 homo sapien
28	334	6.1	596	1 MERL_MOUSE	P46662 mus musculus
29	333.5	6.1	1189	1 PTNE_MOUSE	Q62130 mus musculus
30	327	6.0	572	1 MOES_LYTV	P52962 lytechinus
31	327	6.0	1051	1 YC94_HUMAN	Q92422 homo sapien
32	308	5.6	2485	1 PTND_HUMAN	Q12923 homo sapien
33	302	5.5	578	1 MOES_HUMAN	P46150 drosophila

RESULT 1

ID	4L_HUMAN	STANDARD;	PRT;	864 AA.
AC	P11171; P11176; Q9Y578; Q9Y579; Q14245;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PROTEIN 4.1 (BAND 4.1) (P4.1) (EPB4.1).			
GN	EPB41 OR E41P.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
[1]				
RN	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).			
RP	MEDLINE=88234496; PubMed=3375238;			
RA	Tang T.K., Leto T.L., Correas I., Alonso M.A., Marchesi V.T.,			
RA	Benz E.J. Jr.;			
RT	"Selective expression of an erythroid-specific isoform of protein			
RT	4.1.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 85:3713-3717(1988).			
[2]				
RN	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).			
RP	MEDLINE=89132003; PubMed=3223413;			
RA	Tang T.K., Leto T.L., Marchesi V.T., Benz E.J. Jr.;			
RT	"Expression of specific isoforms of protein 4.1 in erythroid and non-			
RT	erythroid tissues.";			
RL	Adv. Exp. Med. Biol. 241:81-95(1988).			
[3]				
RN	SEQUENCE FROM N.A., AND SEQUENCE OF 378-393 (ERYTHROID ISOFORM).			
RP	MEDLINE=87092279; PubMed=3467321;			
RA	Conboy J., Kan Y.W., Shohet S.B., Mohandas N.;			
RT	"Molecular cloning of protein 4.1, a major structural element of the			
RT	human erythrocyte membrane skeleton.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:9512-9516(1986).			
[4]				
RN	SEQUENCE FROM N.A. (ISOFORM 3).			
RP	MEDLINE=91217063; PubMed=2022644;			
RA	Conboy J.G., Chan J.Y.C., Chasis J.A., Kan Y.W., Mohandas N.;			
RT	"Tissue- and development-specific alternative RNA splicing regulates			
RT	expression of multiple isoforms of erythroid membrane protein 4.1.";			
RL	J. Biol. Chem. 266:8273-8280(1991).			
[5]				
RN	SEQUENCE FROM N.A. (ISOFORM 1).			
RP	Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;			
RT	"Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5.";			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
[6]				
RN	SEQUENCE OF 157-227 FROM N.A., AND VARIANT ILE-214.			
RP	Lichtenauer U., Huang S.C., Vortmeyer A., Zhuang Z.;			
RT	"Valine to isoleucine polymorphism in exon 4 of human protein 4.1.";			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
[7]				
RN	SEQUENCE OF 648-714.			
RP	MEDLINE=87008553; PubMed=3531202;			

34	274	5.0	1062	1 YAL3_HUMAN	Q9Y216 homo sapien
35	3038	4.8	3038	1 TRIO_HUMAN	O75962 homo sapien
36	247.5	4.5	646	1 Y142_HUMAN	Q14155 homo sapien
37	208	3.8	1275	1 GNRP_HUMAN	Q13972 homo sapien
38	206	3.8	1244	1 GNRP_RAT	P28818 rattus norv
39	205	3.8	90	1 4L1_CHICK	P12264 gallus gall
40	205	3.8	519	1 TIM_HUMAN	Q12774 homo sapien
41	197.5	3.6	1262	1 GNRP_MOUSE	P27671 mus musculu
42	193.5	3.5	1919	1 HAPD_RAT	P97924 rattus norv
43	181	3.3	2044	1 SIF2_DROME	P91620 drosophila
44	181	3.3	2064	1 SIF1_DROME	P91621 drosophila
45	178.5	3.3	665	1 YL14_CAEEL	Q11100 caenorhabdi

ALIGNMENTS

RA Correas I., Speicher D.W., Marchesi V.T.;
 RT "Structure of the spectrin-actin binding site of erythrocyte protein
 RL 4.1";
 RL J. Biol. Chem. 261:13362-13366(1986).
 RN [8]
 RP PHOSPHORYLATION AT TYR-660.
 RA MEDLINE=91271361; PubMed=1647028;
 RX Subrahmanyam G., Bertics P.J., Anderson R.A.;
 RA "Phosphorylation of protein 4.1 on tyrosine-418 modulates its
 RT function in vitro";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5222-5226(1991).
 RN [9]
 RP CARBOHYDRATES.
 RA MEDLINE=90036892; PubMed=2808371;
 RX Inaba M., Maede Y.;
 RT "O-N-acetyl-D-glucosamine moiety on discrete peptide of multiple
 RL protein 4.1 isoforms regulated by alternative pathways.";
 RL J. Biol. Chem. 264:18149-18155(1989).
 CC -I- FUNCTION: PROTEIN 4.1 IS A MAJOR STRUCTURAL ELEMENT OF THE
 CC ERYTHROCYTE MEMBRANE SKELETON. IT PLAYS A KEY ROLE IN REGULATING
 CC MEMBRANE PHYSICAL PROPERTIES OF MECHANICAL STABILITY AND
 CC DEFORMABILITY BY STABILIZING SPECTRIN-ACTIN INTERACTION. BINDS
 CC WITH A HIGH AFFINITY TO GLYCOPHORIN AND WITH LOWER AFFINITY TO
 CC BAND III PROTEIN.
 CC -I- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3, ERYTHROID,
 CC NON-ERYTHROID A AND NON-ERYTHROID B; ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -I- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
 CC AND EACH PHOSPHORYLATION EVENT SELECTIVELY MODULATES THE PROTEIN'S
 CC FUNCTIONS.
 CC -I- PTM: PHOSPHORYLATION ON TYR-660 REDUCES THE ABILITY OF 4.1 TO
 CC PROMOTE THE ASSEMBLY OF THE SPECTRIN/ACTIN/4.1 TERNARY COMPLEX.
 CC -I- PTM: O-GLYCOSYL-TERMINAL; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS
 CC IN THE CARBOXYL-TERMINAL DOMAIN.
 CC -I- DISEASE: DEFICIENCY OF E4P1 IS A CAUSE OF HEREDITARY
 CC ELLIPTOCYTOSIS (HE) AND OF HEREDITARY PYROPOIKILOCYTOSIS (HPP).
 CC -I- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; J03796; AAA35793.1; -
 DR EMBL; J03796; AAA35794.1; -
 DR EMBL; M14993; AAA35795.1; -
 DR EMBL; M61733; AAA35797.1; -
 DR EMBL; AF156225; AAD4222.1; -
 DR EMBL; AF156226; AAD4223.1; -
 DR PIR; A60244; MMHUL4.
 DR PIR; A26656; MMHUE4.
 DR GlycosuiteDB; P11171; -
 DR MIM; 130500; -
 DR MIM; 266140; -
 DR InterPro; IPR000299; Band_4.1.
 DR Pfam; PF00373; Band_41; 1.
 DR PRINTS; PR00935; BAND41.
 DR SMART; SM00299; B41; 1.
 DR PROSITE; PS00660; BAND_41_1; 1.
 DR PROSITE; PS00661; BAND_41_2; 1.
 DR PROSITE; PS50057; BAND_41_3; 1.
 DR Structural protein; Alternative splicing; Cytoskeleton;
 KW Phosphorylation; Pyropoikilocytosis; Glycoprotein; Elliptocytosis;
 KW Polymorphism.
 FT DOMAIN 264 420 BAND 4.1-LIKE.
 FT DOMAIN 488 614 SPECTRIN-LIC.
 FT DOMAIN 615 714 SPECTRIN-ACTIN-BINDING.
 FT DOMAIN 715 864 CARBOXYL-TERMINAL.
 FT MOD_RES 660 660 PHOSPHORYLATION (BY EGFR).
 FT VARSPLIC 1 209 MISSING (IN NON-ERYTHROID ISOFORM B).

FT VARSPLIC 228 262 ERYTHROID ISOFORM AND ISOFORM 3).
 FT MISSING (IN NON-ERYTHROID ISOFORM A AND
 FT NON-ERYTHROID ISOFORM B).
 FT VARSPLIC 616 648 MISSING (IN ERYTHROID ISOFORM, NON-
 FT ERYTHROID ISOFORM A, NON-ERYTHROID
 FT ISOFORM B AND ISOFORM 2).
 FT VARSPLIC 635 648 MISSING (IN ISOFORM 3).
 FT VARSPLIC 649 669 MISSING (IN NON-ERYTHROID ISOFORM A AND
 FT NON-ERYTHROID ISOFORM B).
 FT VARSPLIC 772 805 MISSING (IN ERYTHROID ISOFORM).
 FT VARIANT 214 214 V -> I.
 FT /FTIG-VAR-009122.
 FT CONFLICT 51 51 Q -> H (IN REF. 5).
 FT CONFLICT 76 76 S -> N (IN REF. 5).
 FT CONFLICT 168 168 F -> S (IN REF. 6).
 FT CONFLICT 259 259 A -> T (IN REF. 5).
 FT CONFLICT 665 665 N -> S (IN REF. 5).
 FT CONFLICT 679 679 K -> E (IN REF. 5).
 FT CONFLICT 802 802 K -> Q (IN REF. 4 AND 5).
 SQ SEQUENCE 864 AA; 97016 MW; B4731249D7FBE31 CRC64;
 Query Match 13.0%; Score 708.5; DB 1; Length 864;
 Best Local Similarity 32.2%; Pred. No. 4.2e-33;
 Matches 192; Conservative 87; Mismatches 224; Indels 93; Gaps 18;
 QY 16 APENSGISTLERGOKPPPTPSGKLVSIKIQIMLDLTQEAPEVQORAPGKVLDDAVCHNLN 75
 DB 188 SPQSKAETELKASQK--PIKRNHMHCKVSLDDTVYECVVEKHAKGQDLLKRVCEHLN 245
 QY 76 VEGDYFGLFEPFHKKITVWMLDLKPIVKQIRR--PKHVVKVVKVFPFPPDHDTQOEELTRY 134
 DB 246 LEEDYFGLAIDWNATSKTWLDSAKEIKKQVGVWPNFT--FNVKFPDPDPAQLTEDITRY 303
 QY 135 LFALQVKQDLAQLRITCNDSALLSHIVQSEIGDFDEALDR--EHLANKYIPIQQ--DAL 192
 DB 304 YLCQLQRDIVAGLPCSFATLALGYSYIQSELGVDYDPDELHGVVDYDFKLAAPNQTKE 363
 QY 193 EDKIVFEHNNHIGOTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVF 252
 DB 364 EKVYMLHKSYSRMTQAADLEFLNENAKKLSMGVDLHRAKADLEGVDDIILGVCSSGLLYV 423
 QY 253 QGFTKINAFNNAVKRLSPKRRFLIKLPDANSAYODTLEFLMASRDFCKSPWKICVEH 312
 DB 424 KKKLRINRPNPWLKISYKRSFFKIRPGEQOYESTIGFKLPSYRAAKKLWKVCVEH 483
 QY 313 HAFRLFEFPKPKPVLFPSRGSSFRSGRTQKQVLDYVKEGGHKKVQPERKHSKHSIR 372
 DB 484 HTFFRL--TSTDITPKSKFLALGSKFRYSRGTQAOTQASALIDRPAHPHFERTASK--RASR 541
 QY 373 SLASQPTLNSEVLEQSQOSTSLTFCGEAESPGG---QSCRRGKEPKV-----SA 419
 DB 542 SLGAAAVDSADRSRPTSAITAQTQOQVAE--GGVLDASAKKTVPKAKQETVKAEVKKE 599
 QY 420 GEPGSHSPAPRRS-----PAGNKQADGAASAPTEE----- 450
 DB 600 DEPPEQAEPEPTKAVKVEKTHIEVTPTNSGDTQKLAETEDLIRMRKKKRLRDGENI 659
 QY 451 -----EEEV-----VKDRTOOSKPOPPPOPSSTGSLTGSPPH--LSELS 484
 DB 660 YTRHSLMLEDLKXSOEIEKKHHASISELKNFMESVPE--PRPSEWDKRLSTHSPPTLN 718
 QY 485 VNSQ--GGVAP-----ANVTLSNLSPLDTPKQASPLISPLINDQA-----CPRTDD 527
 DB 719 INQIPTGEGPPLVKVTQTVTISDNANAVKSEIPTKDVPIVHTETKTITVEAQTDD 774
 RESULT 2
 ID PTN4 HUMAN
 AC P29074;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)

DR EMBL; AB002336; BAA20796.1; ALT_INIT.
DR EMBL; ALI21895; CAC09920.1; -
DR MM; 602879; -
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
KW Cytoskeleton.
FT DOMAIN 94 308 BAND 4.1-LIKE.
FT CONFLICT 484 484 MISSING (IN REF. 2).
FT CONFLICT 729 729 MISSING (IN REF. 2).
SQ SEQUENCE 881 AA; 98502 MW; D923CF554EDB41D3 CRC64;

Query Match 12.7%; Score 694; DB 1; Length 881;

Best Local Similarity 30.1%; Pred. No. 2.9e-32;
Matches 202; Conservative 96; Mismatches 242; Indels 130; Gaps 22;

QY 18 ENSGIS---TLERGKPPPTPSGKLVS--IKIOMLDDTQEAFAEPORAPGVLLDAVCNH 72
DB 70 EADGLSERTPSKAQSPQIAKYKAICRVILLDASEYCEVEKRGQVFLDVCBH 129
QY 73 LNLVEGDFGLFPPDHKKITVWLDLKPVIKQIR-RPKHVVVVVFVVFPPDHTQOEEL 131
DB 130 LNLLEKDYFGLTFCDADSQKNWLDPSKIKQIRSPWNEA--FTVKFYPPDPAQLTEDI 187
QY 132 TRYLFALQVQDLAOGRLTNCNDSAAILLISHIVQSEIGDFDEALDRH-----LAKNKYIP 187
DB 188 TRYLCQLLRADITGLPCSFVTHALLGSYVAQAEIGDYDA---EHSVGNYSSELRFPAP 244
QY 188 QOD-ALEDKIVFHHNIGOTPAESDFOLLEIARLMYGLRLHPAKDREGTKINLAVAN 246
DB 245 NQTELEERIMELHKTGTGTPGEAEIHFLENAKKLSMYGVDLHAKDSGIDIMLGVCA 304
QY 247 TGLVFOGFTKINAFNWKVRLKFRKRFLKRLPRDANSAYODTLEFLMASRDFCKSF 306
DB 305 NGLLIYRDLRINRFAMPKILKISYKRSNFYKIRPGYEQEFESTIGFKLPNHRSAKRLW 364
QY 307 KICVEHAFRLPEEPKPKPVLFVSGSSFRSGRTQKQVLDYVKEGHHKVKQFERKHS 366
DB 365 KVCIEHHTFRLV-SPEPPKGLV-MGSKFRYSGRTQAQTRQASALIDRAPPFERSSS 422
QY 367 KHSI-----RSLASO-----PTELNSE 384
DB 423 KRYTMSRLDGAERSPASVSENHADPGDGDKRDEDCESGQSRSEAEGEVTRPTKIKEL 482
QY 385 VLQO-----SQOSTSLTFGEAGSPGGQSCRRGKPKVSAGEPGSH 425
DB 483 KPEQETTPRHKBFLDKPEDVLLKHQASINELKRTLKEPNKSLIHRDWRERRLPSSP 542
QY 426 PSAPRSPAGNKQADGAASAPTEEEVEVVKDRTQSKPPQPPSTGS-----LT 475
DB 543 ASPSPKGT---KANERAGLRGSEK---KPPRPAPESDGTGDDQDQDQDRTVFL 594
QY 476 GSPHL-----SELSVNSQGV-APANVTL-----SPNLSPTKQASP-- 511
DB 595 KDNHLATERKCSITVSSSTSLAEVDFTVIGYHGSADFDRSLPELDRKSDSGTEG 654
QY 512 -LISPLNDAQCPRTDDEGRKRFPPTDKAFIAKEVSTERTYTKDLEVIITSWFQSTV 570
DB 655 LLFSRLN-KGAPSQDDSGGIED--SPDRG-----ACSTPDPQPEVPKVTETMTVSSLA 706
QY 571 SKEDAMPEAL 580
DB 707 IRKKIEPAV 716

RESULT 4

41_MOUSE

ID 41_MOUSE

STANDARD;

PRT; 858 AA.

AC P48193;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN 4.1 (BAND 4.1) (P4.1).
GN EPB41 OR EPB41.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=93155238; PubMed=8429050;
RA Huang J.-P., Tang C.-J.C., Kou G.-H., Marchesi V.T., Benz E.J. Jr.,
RA Tang T.K.;
RT Genomic structure of the locus encoding protein 4.1. Structural
RT basis for complex combinatorial patterns of tissue-specific
RT alternative RNA splicing.;
RL J. Biol. Chem. 268:3758-3766(1993).
CC -!- FUNCTION: PROTEIN 4.1 IS A MAJOR STRUCTURAL ELEMENT OF THE
CC ERYTHROCYTE MEMBRANE SKELETON. IT PLAYS A KEY ROLE IN REGULATING
CC MEMBRANE PHYSICAL PROPERTIES OF MECHANICAL STABILITY AND
CC DEFORMABILITY BY STABILIZING SPECTRIN-ACTIN INTERACTION. BINDS
CC WITH A HIGH AFFINITY TO GLYCOPHORIN AND WITH LOWER AFFINITY TO
CC BAND III PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: THE DIVERSITY OF PROTEIN 4.1 ISOFORMS
CC ARISES BY ALTERNATIVE SPLICING
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00919; AAA37123.1; -
CC EMBL; L00919; AAA37122.1; -
CC MGD; MGI:95401; Epb4.1.
CC InterPro; IPR000299; Band_4.1.
CC Pfam; PF00373; Band_41; 1.
CC PRINTS; PR00935; BAND41.
CC SMART; SM00295; B41; 1.
CC PROSITE; PS00660; BAND_41_1; 1.
CC PROSITE; PS00661; BAND_41_2; 1.
CC PROSITE; PS50057; BAND_41_3; 1.
KW Structural protein; Alternative splicing; Cytoskeleton;
KW Phosphorylation.
FT DOMAIN 265 421 BAND 4.1-LIKE.
SQ SEQUENCE 858 AA; 95990 MW; 5F2FEF077946134E CRC64;

Query Match 12.7%; Score 692; DB 1; Length 858;
Best Local Similarity 36.0%; Pred. No. 3.7e-32;
Matches 179; Conservative 65; Mismatches 201; Indels 52; Gaps 12;

QY 5 EQRPTGSRLL---GAPENSGISTLE----RGQKPP-----PTPSGKLVSIIKQLMLDDTQ 51
DB 163 EHREDPDSETKEGETECSEKVEDPESRAEREPEASQKPVRRHNMHCKVSLDDTV 222
QY 52 EAEPVQAPAGPKVLLDAVCNHLNLVEGDFGLFPPDHKKITVWLDLKPVIKQIRR-PKH 110
DB 223 YECVVEKHANLQDLKRVCEHNLLEEDYFGLAWDSATSKTWLDSAKEKKQVGPWN 282
QY 111 VVVKFVVPFPDHTQOEELTRYLFALQVQDLAOGRLTNCNDSAAILLISHIVQSEIGD 170
DB 283 FT--FNWKFVPPDPAQLTEDITRYLCLQRLQDIVAGRLPCSPATLALLGSYTIQSELGD 340
QY 171 FDEAL-DREHLAKNKYIPQO-DALEDKIVFHHNIGOTPAESDFOLLEIARLMYGI 228
DB 341 YDEPHGMDYVSDFKLAPNQTKELEKVMELHKSYSMTPAQADLEFLENAKKLSMGVD 400


```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Lightning J., Thomas K.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 844-950 FROM N.A.
RX MEDLINE=91139172; PubMed=1704870;
RA Matthews R.J., Flores E., Thomas M.L.;
RT "Protein tyrosine phosphatase domains from the protochordate Styela
RT plicata.";
RN Immunogenetics 33:33-41(1991).
RP CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z36237; CAA85272.1; -.
DR EMBL; Z48241; CAA85272.1; JOINED.
DR EMBL; Z48241; CAA88287.1; -.
DR EMBL; Z36237; CAA88287.1; JOINED.
DR EMBL; M38013; AAA28127.1; -.
DR HSSP; P28827; 1RPM.
DR WormPep; C48D5.2; CE17578.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00102; X_phosphatase; 1.
DR PRINTS; PRO0700; PRTPHPPTASE.
DR PRINTS; PRO0935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00194; Ptpc; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00557; BAND_41_3; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Cytoskeleton.
FT DOMAIN 87 245 BAND 4.1-LIKE.
FT DOMAIN 777 1026 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 952 952 BY SIMILARITY.
SQ SEQUENCE 1026 AA; 115093 MW; 2F43F7A614EDBC59 CRC64;

Query Match 11.8%; Score 644; DB 1; Length 1026;
Best Local Similarity 33.2%; Pred. No. 2.6e-29;
Matches 188; Conservative 82; Mismatches 212; Indels 84; Gaps 20;

QY 13 RLCAPENSGISTLERGKPPPTSGKLVSILKIQMLDDTQEAFFVQAPACKVLLDAVCNH 72
Db 2 RLGSNSYDVQRTAIGQTPVKTPPPNQIRCTVTFELDSYHFEIKNSLGVILLEKVFNY 61
QY 73 LNLVEGDYFGLFEP-----PDHKRTVMLDLKPVKQ-IRRPKRHVVKVFFVPPPDH 124
Db 62 LEIENDYFGLVFIADVNSSAQKK---WLDPSKRLRKQMCIPPYHLF--FRVKFVRDP 116
QY 125 TQLOEBELTYFLAQVKQDLAQGRLCNTDTSALLSHIVQSIGDFDE----ALDREHLA 181

```

```

Db 117 NLRDEFTRFQFYQQVRQNLREGRLPCNEGSLALLASYVYVQAEVGFEEKTHGMSRTCLC 176
QY 182 -KNKIYPOQDALEDKIVEPHNHIGOTPAESDFQLEIARRLEMYGIRLHPAKDREGTKI 240
Db 177 YKIQPATLPDDFSRVAELHQLHIGQTPDVAEQNFIDHARRLEMYGMDVYDGVDAHLPI 236
QY 241 NLAVANTGILVFGQFTKINAFNNAKVRKLSFKRRLIKLRPDANSAYQDTLEFLMASRD 300
Db 237 EICGVAGVIGKVPHEGIKMNEYAMVRKLSFKKKQFQV-LVANEDGVSETIMIFNIMSAK 295
QY 301 FCKSFVKICVEHAFPLFEPEPKPVLFSGSSFRFSRGTQKQVL---DYVKEGGHK 357
Db 296 ICKLLMKCCIEQHTFLKTPPK-TPOKVFNFSGFRYSRGTETQYLENEHRKSAGHR 354
QY 358 KVQFERKHSKHSIRSLASQPTLNSEVLQSOQSTSLTEGGEASP-GGSCRRRGKEPK 416
Db 355 --NFRSLSKSSFLRSTFGS---NTQSDSSRY-TNTTTTSPPELSPSSQLLAR---RL 404
QY 417 VSAGEPGSHPSPA-----PRRSAGNKQADGAASAPTEEEVEVVKD 457
Db 405 LSAARHDTSSDALGYASDGVAVCAPLITPLSPRRT--RDYATDESSAPSLRQRLSKE 462
QY 458 R-----TQOS---KPQPPQSTGSLTSPHLSLSV-----NSOGGVA 492
Db 463 AIYGTQESCKDEKSWTPSMACTSTSPGIHASTASVRPVSSGSTPNGASRKANSYSGYG 522
QY 493 PANVTLSPLNLPDTPKQASPLISPLN 518
Db 523 YATQTOOP-----TSTTNASYSPLYN 543

RESULT 7
YF48.HUMAN
ID YF48.HUMAN STANDARD; PRT; 732 AA.
AC Q9HCM4; Q9H975;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA1548.
GN KIAA1548.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 188-732 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirokawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RL code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK023019; BAB14360.1; -.

```

```

EMBL; AB046768; BABI3374.1; .
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00507; BAND_41_3; 1.
DR Hypothetical protein; Cytoskeleton.
FW DOMAIN 40 255 BAND 4.1-LIKE.
FT CONFLICT 669 732
FT FT
FT TSSGMLAGCEMLLTGTKEGHNGKDGISLISPAPFLVDVAV
FT TSSGPILAEAVLKQKCLLTTEL -> LMSHFGRRSCPEAE
FT VFTDH (IN REF. 1).
SQ SEQUENCE 732 AA; 81758 MW; 76D5BD8CE099E761 CRC64;

Query Match      11.5%; Score 626; DB 1; Length 732;
Best Local Similarity 30.18; Pred No. 1.7e-28;
Matches 191; Conservative 92; Mismatches 221; Indels 130; Gaps 22;

Qy 31 PPPTPSGLVKIQIMLDDDTQEAFEVQRAPGKVLLDAVCNHLNLVEGDYEGFLFPDHHK 90
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 34 PAAGDSKSIICTIRSVSLDGTSDVDLPKKAKGQELFDQIMVHLDLIESDYGLRFMDSAQ 93
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 91 ITVWLDDLKPVIQIRRPKHVVVFVKKFPPDPDTQOEELTRYLFALQVKKDQAQRLT 150
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 94 VAHVLDDGTSIKKGKIGSPCYCLHRVKFYSEPNLNREELTRYLVFLQLQKLIDLSGLD 153
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 151 CNDTSAALLSHIVQSEIGDFDEALDREH----LAKNKYIP-QQDALEDKIVEFHNNHG 205
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 154 CPFDFAVOLAAINYLAQELGYDLA--EHSPELVSEFRVPQTQEEMLATIFEKWKEYRG 210
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 206 QTPADEPQLLEIARRLEMVGIRLHKADREGTKINLAVANTGILVFOGTTKINAFNAWK 265
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 211 QTPAQAEYNLYNAKAWLEMVGMHVYKARGNDYSGLTGTGVLPESDITKIGLFFWPK 270
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 266 VRKLSFKRKR-F-LIKLRPDANSAYOD-TLEFLMASRDCKFWKICVBHHFAFFRLFEEP- 322
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 271 ITRLDFKNKTLTVVVEDDDQGKEQHTFVRLDHPKACKHLWKCAVEHHFAFFRL-RGVP 329
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 323 -KPKPKPVLSRGSFSRFGRTQOVL DYKVEGGHKVQFERKHSKHISRSL-----AS 376
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 330 QKSHRSFGIFLGSRFRYSGTEYOTTKNK--ARRSTSFERRRSGSKRYSRRTLQMKACAT 387
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 377 OPTLENSEVLQSQQSTSLTLEGEGASPGQSCRGRKPKVSPGSGHPSPAPRRSPAG 436
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 388 KPEEL-----SVHNNVYSTQSGQAWGMRSA LPVSPSIS----- 422
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 437 NKQADGAASAATEEEEEVVKORTQOSKQPPO-----PSTGSLTGSPHLSLSVNS 487
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 423 -----SAPVPVEIE-----NLQSPGTDQHDKRKICPLNIDLNLSPDLLTETI-- 464
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 488 OGVAPANVTLS-----PNLS-----PDTKA-----SPLISPL 516
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 465 -GDVIGASDTHMETSQLNDVNVAITRLPGLGEPEVEYETLKDTSERKLKOLEMENSPLISPR 523
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 517 LN-----DQACPRTDEDGRR-KRPPTD-----KAYPIAKEVSTERTY 555
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 524 SNIDWNINSQEVEVKLTBKCLNNVIESGLNVNMVRPPDFKSNILKAQEAHVKKYTKEDSL 583
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Qy 556 L--KDLEVITSWFSTVSKEDAMP---EALKSLI 584
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 584 LSHKANVODAATNASAVLNENNVLPKRESLETLM 617
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 8
PTN3_HUMAN PTN3_HUMAN STANDARD; PRT; 913 AA.
AC P26045;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48);
```

```
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.  
KW Structural protein; Cytoskeleton; Hydrolase.  
FT DOMAIN 84 241 BAND 4.1-LIKE.  
FT DOMAIN 510 582 PDZ.  
FT DOMAIN 670 913 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT_SITE 842 BY SIMILARITY.  
SQ SEQUENCE 913 AA; 104029 MW; 29A539ACDE2F1515 CRC64;  
  
Query Match 9.9%; Score 541; DB 1; Length 913;  
Best Local Similarity 24.3%; Pred. No. 1.7e-23;  
Matches 233; Conservative 150; Mismatches 346; Indels 228; Gaps 41;  
  
QY 27 RGQPPPTPGSKLISIRIOMLDQAEFVQAPAGKVLDDAVCNHNLNVEGDFGLLEFP 86  
DB 16 RTSELPREKTRSEVICSIFHFDGVQFVKYQDGTGVLLDMVHNLGVTEKEVFLGHD 75  
QY 87 DHKITY-WDLLKPIYKQIRRPKHVVVVFPPDPHTQLOEELTRYLFALQVKODLA 145  
DB 76 DSDVSPRWLEASKPIRKQLKGGPPCTLHFRVRFPIPDNTLQEQTRHLYFLQLKMDIC 135  
QY 146 OGRITCNDTSAALLSHIVQSEIGDFDEALDRE-HLAKNKYIPQDALED---KIVEFHH 201  
DB 136 EGRITCPLNSAVVLASAVOSHFGDYNSSITHPGYLSHDFIPDN--EDFLTKVESLHE 193  
QY 202 NHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGGFTKINAF 261  
DB 194 QHSLGKQSEAEVCINARTLDYFVGLHSGRDLNLDLMIGTASAGVAYRYKICTSFY 253  
QY 262 NNAVKRLSKPKRFLIKLRPDANSVQDTLEFLMASRDFCKFWKTCVBHHAFFRLLEE 321  
DB 254 PWNILKISFKRKKFFTHQKQAESREHIVAFNMLNYSKCNLWKSCEVHHHTFFQA-KK 312  
QY 322 PKPKPKVL---PSRGSFSGRTQKQLD-VYKE---GG-----HKVQPERKHSKI 368  
DB 313 LLPQEKVNLVSOYWTMGR-----NTRKSVNNQYCKKVIIGVGNWNPAMRRSLSEVHLETK- 366  
QY 369 HSTRSLASQPTELNSELVLEOQSOSTSLTFGEAESPGQSCRRGKPKVAGEPGGHPSP 428  
DB 367 ----SLPSRPPI-----TPNWR-----PRL-----RHEIR 389  
QY 429 APRSPAGNKQADGAASAPT--BEERVVKDRTQOSKQPQPQSTGLTSGPHLSLSVN 486  
DB 390 KPHRS-----SADLANEMTYITETEDVFTYKGLAPO--DSDSEVSQNRSPHQSLEN 443  
QY 487 SQGVAPANTVLSPLSPDTKQASPLISPLNDQACPTDDEGRKRKREPTDKAYFIAK 546  
DB 444 N-----PAQSYL-----TQKSSSVSPSSNAPGSCSPGVQQL-----LDDEHRTVK 486  
QY 547 EVSTERTYLLKLEVTISWFQSTVSKEDAMPEAKSL-IPNFEPLHKHFTNFKLEIQR 605  
DB 487 GGSTEDAS-----QYQCDKNDGSDVLLVLRITPDED--GKGFNLKGGVDQK 532  
QY 606 LALWEGRSNAQ-----IRDYQIRGDVNLKNIQGMK-----LAAHLWKHSLEALEA 650  
DB 533 MPLVVRINPESPADTICPLKNEGDQVLVINGRDISHTHDQVVMFTKASRESHSREL-A 591  
QY 651 LENGIKSRLENFCRDFELQKCYLPLNTFLRLHRLMHYKQVLERLCKHHPHSHADF 710  
DB 592 LVIRRAVRFADEKSEDELNL-----PPEAFPMC-----PEGGD- 628  
QY 711 RDCRAALAEITEMVAOLHGTMIKME-----NFQKLHELKLDLIGDNLVWPGREF 760  
DB 629 -----TLEGSMQALKKLGESGTVLIOFEQLYRKK-----PGLAI 662  
QY 761 IRLGSLKSLSGKLGQORMFLFNDVLLY-TSRGLTASNQKRVHG-----QLPLYGMTIEE 814  
DB 663 -----TFALKPQNLKDN-----YKDVLPDYTRVLLOGNEDYINASYNNMVEIPAANLVNKY 714  
QY 815 SEDEWGVPH-CL-----TLRGQROSIIVASSRSEMEKRWEDIQMAIDLAEKSSSAPPEFL 869  
DB 715 IATGGPLPHTCAQFQWVWVDQKLSLIV-----MLTTLTERGRTCKHQYV 758
```

```
QY 870 ASSPPD-----NKSPDEATAAQDESDDLASRTSLERQAPHRGNTMVHVWCW 916  
DB 759 -PDPPDVNHHGFIHQSESDCTIAYVSRMLVNTQTGEHTVTH-----LQYVAV 809  
  
RESULT 9  
NBL4_BRARE  
ID NBL4_BRARE STANDARD; PRT; 619 AA.  
AC O57457;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE BAND 4.1-LIKE PROTEIN 4 (NBL4 PROTEIN).  
GN NBL4.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
OX NCBI_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98212541; Pubmed=9551184;  
RA Kelly G.M., Reversade B.;  
RT "Characterization of a cDNA encoding a novel band 4.1-like protein in  
RT zebrafish";  
RL Biochem. Cell Biol. 75:623-632(1997).  
CC !- FUNCTION: NOT KNOWN, BINDS CALMODULIN.  
CC !- TISSUE SPECIFICITY: IN ADULTS, IT IS FOUND IN THE OVARY, EYE,  
CC HEART, AND BRAIN, BUT NOT IN GUT OR SKELETAL MUSCLE.  
CC !- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AF025306; AAB97965.1; -.  
DR ZFIN; ZDB-GENE-990415-20; nbl4.  
DR InterPro; IPR000299; Band_4.1.  
DR Pfam; PF00373; Band_41; 1.  
DR PRINTS; PR00935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR PROSITE; PS00660; BAND_41_1; 1.  
DR PROSITE; PS00661; BAND_41_2; 1.  
DR PROSITE; PS00057; BAND_41_3; 1.  
KW Structural protein; Cytoskeleton.  
FT DOMAIN 70 230 BAND 4.1-LIKE.  
SQ SEQUENCE 619 AA; 70708 MW; 0DFD7324DE513663 CRC64;  
  
Query Match 9.7%; Score 532; DB 1; Length 619;  
Best Local Similarity 29.4%; Pred. No. 3.1e-23;  
Matches 146; Conservative 83; Mismatches 178; Indels 90; Gaps 14;  
  
QY 43 KIQMLDD-----TQAEFVQAPAGKVLDDAVCNHNLNVEGDFGLLEFPDHKKITVWLDL 97  
DB 14 EVLLDESKLLLTQOQGIKKSTRGSVLDYVFSHVNLAETEYFGVRYCDRSHQTFWLDP 73  
QY 98 LKPIVKQ---IRRPKHVVVVFVFPDPHTQLOEELTRYLFALQVKQDLAQGRITCNDT 154  
DB 74 SKTLAEHKDLIATGPPTLYFGVKFYAEDDPGLKEITRYOFFLQVQKQDLQGRILPCAFN 133  
QY 155 SAALLSHIVQSEIGDFDEALDRE-HLAKNKYIP-QDALEDKIVEFHHNHIGQTPAESD 212  
DB 134 ISQAALAAIYSELGSDYDPYKHTAGVSEYRFPVDFQKEDLEDSEIQHTLLQGVPAEAE 193  
QY 213 FQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGGFTKINAFNNAVKRLSKFK 272  
DB 194 NNLATAKLTLEMVGVLDLHPVGEKQAEYFLGLTPGVGVVYVKNKTQVCKYFWPRITKYVFK 253
```



```
QY 273 RRRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFRLFEPEK---PKPKPV 329
Db 254 ETQF--ELRVLRGDCNETSFFFDAAACKTACKNLWCCVEHHTFFFRMPENESLTKRLSK 311
QY 330 LFSRGSFRSFGRTQKV-LDVVKEGGHKKVQFERKHSIHISIRSLASOPTELN----- 382
Db 312 FSSLSGKHYRSGKTAMQIGRESTEVLPLRADLQVIRTSKTYPKRS--TOPAGRNNGQAV 369
QY 383 -----SEVLEQSQSTSLTFGGAESPG--GOSCRGKGP 415
Db 370 TKMNTSEGPKTSALTVPKSPRVKAESASVQOEKPSAPWEADAPQSGLYNSASERNKSP 429
QY 416 KVSAGEPGSHSPAPRSPAG-----NKO-----ADGAAS 445
Db 430 KP-----PKAH-----RRSPSGGSENEPHRRGQTADDAQANKOHRRRRSRGRNTSSGSES 480
QY 446 APTEEEEVVKDRTOQS 462
Db 481 ENSNREHKKRNRQS 497

RESULT 10
NBL4_HUMAN
ID NBL4_HUMAN STANDARD; PRT; 598 AA.
AC Q9HCS5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE BAND 4.1-LIKE PROTEIN 4 (NBL4 PROTEIN).
GN EPB414.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334634; PubMed=10874211;
RA Ishiguro H., Furukawa Y., Daigo Y., Miyoshi Y., Nagasawa Y.,
RA Nishiwaki T., Kawasoe T., Fujita M., Satoh S., Miwa N., Fujii Y.,
RA Nakamura Y.;
RT "Isolation and characterization of human NBL4, a gene involved in the
RT beta-catenin/tcf signaling pathway."
RL Jpn. J. Cancer Res. 91:597-603(2000).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. HIGH LEVELS OF
CC EXPRESSION IN BRAIN, LIVER, THYMUS AND PERIPHERAL BLOOD LEUCOCYTES
CC AND LOW LEVELS OF EXPRESSION IN HEART, KIDNEY, TESTIS AND COLON.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match 9.5%; Score 518.5; DB 1; Length 598;
Best Local Similarity 25.2%; Pred No. 1.8e-22;
Matches 167; Conservative 123; Mismatches 255; Indels 117; Gaps 22;

```
QY 43 KIOMLDDTOEAEPVOR-----APGKVLLDAYCNHLNVEGDYFGLFEFFDHKKITVNDL 97
Db 14 EYLLDSEKLTLTTOQGIKSTKSGVLDHVFPHNVEIDYFGLRYCDRSQHTWLDP 73
QY 98 LKPIVKQ---IRRPKHVVVYKVFVPPDHTOLOEBELTRYLFALQVKQDLAQGRJTCNDT 154
Db 74 AKTLAEHKEILNTGPPYTYLFGIKFYAEDPCKLKEITRYQFFLOWQDVQGRGPCVN 133
QY 155 SAALISHIVQSEIGDFDEALDRE-HLAKNKYIP-QDDALEDKIVEFHNNHIGQTPAES 212
Db 134 TAAQLGAVAIQSELGSDYDPYKHTAGYVSEYFVDPQKEELEAEIRHKTLMGQIPSEAE 193
QY 213 FOLLETARLEWYGIRLHPAKDREGTKINLAVANTGILYFOGFTKINAFNAWVKRLSEK 272
Db 194 LNYLTRAKSLEMYGVDLHPVYGENKSEYFLGTFVGVVYKNNKKQVGFYFPRITVKHK 253
QY 273 RKRFLIK-LRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFRLFEPEKPKPVLF 331
Db 254 ETQFELRVLGKOCN---ETSFEEARSKTACKHLWKCSVEHHTFFFRMPENESLSRKLS 310
QY 332 SRGS---SFRSGRTQKV-LDYVKEGGHKKVQFERKHSIHISIRSLASOPTELNSEVLE 387
Db 311 KFGSIRYKHRYSGRTALQMSRDLISQLPRPDQNVTRSRKTYPKRIAQTQPAESNT--- 366
QY 388 SQQOSTSLTFGGAESPGSGSCRRGKPKVSGEPGSHSPAPRSPAGNKQADGAASAP 447
Db 367 ISRITANM---ENGNEG-----TIKIIA-----PSPKSFKKAKNE-----NSP 403
QY 448 TEEEEEVVKDRTOQSKPQPQPPSTGSLTGSPLHSELVSNQSGV--APANVTLSPLNLPD 505
Db 404 DTQRSKSLMHSWENGP-----QSGLYNSPDRTKSPKF-PY 439
QY 506 TKQASP-----LISPLNDQACPTDDEDEGRKRKFTDKAYFAKEVSTTERTYTKD 558
Db 440 TRRNPSGSDNDSVQPVRRRKAHNSGEDSDLKQRRRSRRCNTSSGSESENREHKK 499
QY 559 LEVITSWFQSTVSKEDAMPEALKSLIFNPEPLKFTNFKLEIQRLALWEGRSNAQIR 618
Db 500 -----NRRIRQENDMVDSA-----PQWEAV-----LRQEKKNQADPNRSRHR 539
QY 619 DYQIRGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSSRRLENCFQFELQVKCYLPL 678
Db 540 SRSRSDIOAKE-----ELWKHIQELVDPGLSEE-----QLAKEIPTKI 580
QY 679 NT 680
Db 581 ET 582

RESULT 11
NBL4_MOUSE
ID NBL4_MOUSE STANDARD; PRT; 554 AA.
AC P52963;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BAND 4.1-LIKE PROTEIN 4 (NBL4 PROTEIN).
GN EPB414 OR EPB4.114.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95074267; PubMed=7983158;
RA Takeuchi K., Kawashima A., Nagafuchi A., Tsukita S.;
RT "Structural diversity of band 4.1 superfamily members."
RL J. Cell Sci. 107:1921-1928(1994).
CC -!- TISSUE SPECIFICITY: BRAIN, HEART, LUNG, LIVER AND SPLEEN. NOT
CC DETECTED IN THYMUS AND KIDNEY.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; D28818; BAA05978.1; -
CC MGD; MGI:103007; Ebb4.114.
CC InterPro: IPR000299; Band_4.1.
CC Pfam: PF00373; Band_41; 1.
CC PRINTS; PR00935; BAND41.
CC SMART; SM00295; B41; 1.
CC PROSITE; PS00660; BAND_41.1; 1.
CC PROSITE; PS00661; BAND_41.2; 1.
CC PROSITE; PS50057; BAND_41.3; 1.
KW Structural protein; Cytoskeleton.
FT DOMAIN 8 231 BAND 4.1-LIKE.
SQ SEQUENCE 554 AA; 64082 MW; 7E63BAD088DBEBE CRC64;

Query Match 8.7%; Score 475.5; DB 1; Length 554;
Best Local Similarity 30.3%; Pred. No. 4.5e-20;
Matches 135; Conservative 79; Mismatches 182; Indels 49; Gaps 14;
QY 43 KIQMLDDTQFAFEVQR-----APGKVLDAVCNHLNLVGDYFGLFFPDHKKITVWLDL 97
Db 14 EVLLDESKLTLTQQGKIKKSGSVLDVHFRHNLVEIDYFGLRYCDRSHQTYWLDP 73
QY 98 LKPIVQ---IRPKHVVFVVFPPDHTQLOELTRYLFALQVKDLAQRGLRCNDT 154
Db 74 AKTAEHKEKELINTGPPYTLFYGKFAEDPCKLEITRYOFFLQVKDALQRLPCPVN 133
QY 155 SAALLSHVQSEIGDFDEALDRE-HLAKNKYIP-QDDALEDKIVEFHNNHIGOTPAESD 212
Db 134 IAQMGAIALQELGDHDPKHTAGVSEYRFPDQKEELEAEIERIHKTLMGQAFSEAE 193
QY 213 FOLLEIARRLEMGIIRLHPAKDREGTKINIAVANTGILVFQGTKNIAFNWAKRKLSEK 272
Db 194 LNLVLRATKSLMGVLDLHPYVGENKSEYFLGLTPSGVVVYKKNQGVKYFWPRITKVHF 253
QY 273 KRFLIK-LRPDANSAYQDTLEFLMSARDCKSFVKICVCHHAFFRLFEPEKPKPVLF 331
Db 254 ETQFELVLVKDCN---ETSFEEFARSKTACKHLMKCSVEHHFTFRMPDPTESNLRKLS 310
QY 332 SRGS-SFREGRTQKOV-LDYVVEGGHKVQFERKSHKSHIRSLASOPTELNSEVLEQS 389
Db 311 KFGSISYKHYRTALQMSRDLSQLPRPNQNVRSRSKYPKVAOTQPTGSNN--INRI 368
QY 390 QOSTSLTFFRG-----AESPGQSCRRGKPKVYSAGEPGSHPSAPRRSPAGNKQADGAA 444
Db 369 TANTENGENGTKIITAPSP-VKSFKKAK-----NENSPDPQRS-----KS 408
QY 445 SAPTEEE-----EVVVKDRTQSK 463
Db 409 HAPWEENGPGQGLYNSSSDRTKSPK 433

RESULT 12
ID FGD1_MOUSE
AC P52734;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE {FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG}.
GN FGD1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=96081343; PubMed=8535076;
RA Pasteris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,
RA Gorski J.L.;
RT Cloning and regional localization of the mouse facio-genital
RT dysplasia (Fgdl) gene.;
RL Mamm. Genome 6:658-661(1995).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GDP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOG DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FIVE-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U22325; AAA96001.1; -
DR MGD; MGI:104566; Fgdl.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000822; Znf-C2H2.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 2.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS50178; ZF_FYVE; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; Zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 589 688 PH 1.
FT ZN_FING 729 789 FYVE-TYPE.
FT DOMAIN 820 920 PH 2.
SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 8.2%; Score 450.5; DB 1; Length 960;
Best Local Similarity 22.9%; Pred. No. 2.6e-18;
Matches 189; Conservative 131; Mismatches 338; Indels 169; Gaps 29;
QY 320 EEPKPKPVLFSGSSFRSGRTQKQVLDYVVEGGHKVQFERKSHKSHIRSL----- 374
Db 142 EIPGPRSP-----LKRACPKPQVPKPSYLOMPVLPPPEP 179
QY 375 ----ASQPTELNSEV-----LEQQSTSLTFFGEGASPGQSCRRGKPKVSAG----- 420
Db 180 IPPPPSRPLPADPRVAKGLVPRAEASTS-----SAVSSLLIEKPER--EPVIVASDRPAP 233
QY 421 -----EPGSHPSAPRRSPAGNKQADGAAS-----APTEEEEVVKDRTQ-----SKP 464
Db 234 PCVPPEPAMLPOPPQ--PTGSQLPGEASRCLFLAPGPRGGEKVPNRRSDSIDISSP 291
QY 465 QPPQPSGTSLTSGPHLSVNSQGVAPANVTLSPLNSPDTKQASPLSLPLNDQACPR 524
Db 292 SNEETCFVSDGPPHSLCPG-----PALASMPVALADPHRPGSQEVDSDLEEE----- 342
QY 525 TDEDEGRKRFPPT-----DKAYFIAKEVSTERTYLLKDLLEVITSWFQSTV 570
Db 343 EEEEEKEEREIPVPPMERQESVELTVQKVFHIANELLQTEKAYVSRLLHLLDQVFCARL 402
QY 571 SKE-----DAMPEALKSLIFNPEPLKFTNFP-LKEIEQRLALWEGRSNAQIRDYQRIGD 625
Db 403 LEEARNRSFPADVVHGVIFSNICSYICFHQQFLLPELEKRMEDW-----RYPRIGD 454

QY 626 VMLKNTQGMKHLAAHLKWKHSEALENGIKSSRRLENFCRDFELQKVC-YLPNTLTLR 684
 Db 455 ILQKLAPFLKMYGEYKFNFRVAVELVNTWTSTQPKVIIHEVQKEACRNLTLOHHMLE 514
 QY 685 PLRLMHYQVRLERLCKHHPSPSHADFCRAALAEITEMVAQLHGTMIKMFNOKHLK 744
 Db 515 PVQIRPYELLKLDYLLKLPKPHGSPDSKDAKSLLELTATAEHSHNAIRKMERHKLKVV 574
 QY 745 KOLIGDNLVVPREFIRIGLSKLSKSGK--LQORMFFLNDVLLYTSRGLTASNO-FKV 801
 Db 575 ELLGGEEDIVSPTELKELKEGHLKLSAKNKTQDRVLLILFNDRLLYCVPLRLGQKFTV 634
 QY 802 HGOPLPLYGMTIESEDEWGVPHCLTLRGQRSIIIVASRSSEMEKVVEDIQMAIDLAES 861
 Db 635 RARIDVGMELKSSN-LNMPRTFLVSGKRSLEQLQARTEEKKDWQVAINSTL-LKHEQ 692
 QY 862 SSPAPEFLAS-----SPDNKSPD---EATAADOESDDL---SASRTSLERQAPH- 906
 Db 693 TLETFKLLNSTNRDDETPPNSPNVDLGRKAPPIREKEVTMCMRQEPFNSITKRRHHC 752
 QY 907 -----RGNTMVHVCW-----HRNTSVSMV 925
 Db 753 KACGHVVCGKSEFARLIYDNNKSNRVCTDCVVALHGAPGSSPACSOHTPQRRRSILEK 812
 QY 926 DFSIAVENQLSGNLLKRF-KNSNGWOKLWVFTN---FCLFFYKSHQDNHPLASLPLGY 981
 Db 813 QASVAEANSVICSFLHYMEKGGKGWKAFVVPENEPVLYIYCAPQDVKAQRSPLIGF 872
 QY 982 SLTIPSESENIQDYVKLHFKSHV-YFRAESEYTFERMEVIRSA 1027
 Db 873 EVGPPEAGERDRRHVFKI-TQSHLSWYSPETEELQRRMVAVLGRA 918

RESULT 13
 FGDI_HUMAN
 ID FGDI_HUMAN STANDARD; PRT; 961 AA.
 AC P98174;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
 DE (FACIOGENITAL DYSPLASIA PROTEIN).
 GN FGDI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Craniofacial;
 RX MEDLINE=95042764; PubMed=7954831;
 RA Pasternis N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,
 RA Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.;
 RT "Isolation and characterization of the facio-genital dysplasia
 RT (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide
 RT exchange factor.";
 RL Cell 79:669-678(1994).
 CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
 CC BY EXCHANGING BOUND GDP FOR FREE GTP.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA, AND
 CC LESS IN PANCREAS AND LIVER.
 CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE
 CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT
 CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.
 CC -!- SIMILARITY: CONTAINS 1 DEL-HOMOLOGY DOMAIN (DH).
 CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 FIVE-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U11690; AAA57004.1; -
 DR MIM; 305400; -
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; RhoGEF.
 DR InterPro; IPR000822; Znf-C2H2.
 DR InterPro; IPR000306; Znf_FYVE.
 DR Pfam; PF01363; FYVE; 1.
 DR Pfam; PF00169; PH; 2.
 DR Pfam; PF00621; RhoGEF; 1.
 DR SMART; SM00064; FYVE; 1.
 DR SMART; SM00233; PH; 2.
 DR SMART; SM00325; RhoGEF; 1.
 DR PROSITE; PS0178; ZF_FYVE; 1.
 DR PROSITE; PS50003; PH_DOMAIN; 2.
 DR Guanine-nucleotide releasing factor; zinc-finger.
 FT DOMAIN 7 330 PRO-RICH.
 FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
 FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
 FT DOMAIN 590 689 PH 1.
 FT ZN_FING 730 790 FIVE-TYPE.
 FT DOMAIN 821 921 PH 2.
 SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match 8.1%; Score 441.5; DB 1; Length 961;
 Best Local Similarity 23.7%; Pred. No. 8.6e-18;
 Matches 173; Conservative 112; Mismatches 299; Indels 147; Gaps 26;

QY 413 KEPKVSAGE---PGSHPS-----PAPRRSPAGNQADGAAS-----APTEEEEEVVK 456
 Db 220 REPVIASDRPVGCPSPGPPVNLQPTSQPPVPQLPEGEASRCLEFLAPGRDGEKVP 279
 QY 457 DRTQQ---SKPQQPQ-----PSTGLTSGPSHLSSELSVNSOGGVAPANVTLSPNL 502
 Db 280 NRDSGDISISSPSNSETCFVSDGPPSHSLCPGPAL-----ASVPVALADPH 328
 QY 503 SPDTKQASPLISLLNDQACPRTDEDEGRK--RPT-----DKAYFIK 546
 Db 329 RPSQEV-----DSLEEDDEDEEEKREIPVPLMERQESVELTVQKRVFHIAN 379
 QY 547 EVSTTERTYKLEVITSMFQSVSKP-----DAMPAELKSLIFPNFPLHKFTNF-LKE 601
 Db 380 ELLQTEKAYVSRLLHLLDQVFCARLLEARNRSFPADVVHGFISNCSIVCFHQQLLPE 439
 QY 602 IEQRLALWEGRSNAQIRDYQIRIGDVMLKNTQGMKHLAAHLKWKHSEALENGIKSSRRL 661
 Db 440 LKRMSEWD-----RYPRIGDILQKLAPFLKMYGEYKFNFRVAVELVNTWTSTQF 491
 QY 662 ENCRFPELOKVC-YLPNTLTLRPLRLMHYKQVLERLCKHHPSPSHADFCRAALAEI 720
 Db 492 KVTIHEVQKEACNLTQHHMLPQVQIRPYELLKLDYLLKLPKPHGSPDSKDAKSLLEI 551
 QY 721 TEMVAQLHGTMIKMFNOKHLKLDIGDNLVVPREFIRIGLSKLSKSGK--LQORM 778
 Db 552 ATAEHSNAAIRKMERHKLKLVYELLGGEEDIVSPTELKELKEGHLKLSAKNKTQDRY 611
 QY 779 FFLFNDVLLYTSRGLTASNO-FKVHGOPLPLYGMTIESEDEWGVPHCLTLRGQRSIIIV 837
 Db 612 LILFNDRLLYCVPLRLGQKFSVRARIDVDGMELKSSN-LNLPRTFLVSGKRSLELQ 670
 QY 838 ASSRSEMEKVVEDIQMAIDLAESKSSSPAPEFLAS-----SPDNKSPD---EATAADQ 887
 Db 671 ARTEEEKKDWQVAINSTL-LKHEQTLTFTKLLNSTNRDDETPPNSPNVDLGRKAPPIR 729
 QY 888 ESEDDL---SASRTSLERQAPH-----RGNTMVHVCW--- 916
 Db 730 EKEVTMCMRQEPFNSITKRRHCKACGHVVCGKSEFARLIYDNNKSNRVCTDCVVAL 789

```
QY 917 -----HRNTSVSMVDFSIIVNOLSGNLLRKF-KNSNGMQLWVFTN-- 958
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 790 HGVGSSPACSQHTPQRRIISLEKQASVAENSVCISFLHYMEKGGKGWKAFTVVPENE 849
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 959 -FCFLFFKSHQDNHPLASLPLGLYSITPSENIQKDYVFLHFKSHV-YFRASEYX 1016
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 850 PLVLYIGAPQDVKAQSLPLIGFEVGPPEAGERDRRHVKFI-TQSHLSWYFSPETEEL 908
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1017 FERWMEVIRSA 1027
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 909 QREWMAVLGRA 919
      | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14
PTNL_RAT PTNL_RAT STANDARD; PRT; 1175 AA.
AC Q62728; Q62732;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE 2E).
GN PTPN21 OR PTP2E.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=95104449; PubMed=7805871;
RA L'Abbe D., Banville D., Tong Y., Stocco R., Masson S., Ma S.,
RA Fantus G., Shen S.H.;
RT Identification of a novel protein tyrosine phosphatase with sequence
RT homology to the cytoskeletal proteins of the band 4.1 family.;
RL FEBS Lett. 356:351-356(1994).
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + PHOSPHATE.
CC -1- TISSUE SPECIFICITY: PARTICULARLY ABUNDANTLY IN ADRENAL GLANDS.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U17971; AAA62153.1; -
DR EMBL; U18293; AAA62154.1; -
DR HSSP; P18031; 1PTT.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPHTASE.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00057; BAND_41_3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
KW Structural protein; Cytoskeleton; Hydrolase; Alternative splicing.
FT DOMAIN 77 241 BAND 4.1-LIKE.
FT DOMAIN 922 1175 PROTEIN-TYROSINE PHOSPHATASE.
```

```
FT ACT_SITE 1109 1109 BY SIMILARITY.
FT VARSPLIC 1 839 MISSING (IN ISOFORM 2E).
SQ SEQUENCE 1175 AA; 133411 MW; 82A684F1C0F5ECF7 CRC64;

Query Match 7.9%; Score 434; DB 1; Length 1175;
Best Local Similarity 21.8%; Pred. No. 3,1e-17;
Matches 258; Conservative 171; Mismatches 458; Indels 298; Gaps 49;

QY 25 LERQKPPPTSGKIVSIKIQLDDTQEAFAFVQAPQAPVLLDAVCNHLNVLVEGDYFGL 84
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 9 LKTRRTYTVSSKSLVA-RIQLLNNEFVFTLSVESTGQESLEAVAQRLRELTITYFSLW 67
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 85 EPDHKKITVWLDLLKPIVKQIRPKHV--VKFVKFPPDHTQLOEELTYFLFALQVK 141
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 68 YYNQNRWVDLEKPLKKQL--DKHALEPTVYGVVFPVSVSQLOEELTYFLYQLK 125
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 142 QDLAQGLTCNDTSAALLISHIVQSEIGDFDEALDREHLAKNKYIP-----QDALED-- 194
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 126 KQVLEGNLPCTLEQAQLAGLAVQADFGDFQYESQDFLQKFALLPVGMQLDEKLLEEA 185
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 195 -KIVEPHHNIGOTPAESDFOLLEIARRLEMYGIRLHPAKDRGTTKINLAVANTGILV-- 251
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 186 QKVALLHOKYRGLTAPEAEMLYNQEVERMDGYEESYPAKDSQSDISIGACLDGIFVKH 245
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 252 FQGETKINAFNNAKVKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVE 311
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 246 KNGRPV-VFRHDIANMHNKSSFFALEL---ANK--ETIQOTEDMETAKYVWLICVA 299
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 312 HHAFRL-----FEPPKPK-----PKPVLSFGSSFRFSGRTQ 344
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 300 RHKFYRLNQCNIQTQAATLNSVRRGSSSRMSLPKPKQPYAMPPPQLHYNGHYTEFPASSQ 359
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 345 KOVLVYKGG--GHKKVQFER-----KHSKIHISRS-----LASQP 378
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 360 DNVEFVNKNGFYCHSQTSLDRTQIDLSGRIRNGSVYSAHSTNSLTNPQPYLPQSPMSNP 419
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 379 TELNSVLE---QSOOSTSLTFGEAGSPGQS---CRRG-----412
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 420 SIPIGSDVMPDPIPSHRHSALIPPSYRPTPDYESVMKRLNRGMVHADRHSHSLRNIGS 479
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 413 -----KEPKVSAGEGSHP-----SPAPRRSPAGNKQADGAASA 446
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 480 SVAYSRPDALVYSQPIREHPHLASQSAHYFPNLNVSFHSQAPYVPVRRPVVGAVSV 539
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 447 PTEEEEV-VKD-----RTQSKQPQPOPSTGSLTSPHLS-ELSVNSQGVAVANTV 497
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 540 PELTNVQLAQADYPAPNIMRTQVYRPPYPYPRPANSTPDLRHLIYISSN---PDLIT 596
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 498 LSPNLSPTDKQASPL-ISPLLNDOACPRTDDEDEGRKRKRPDTPDKAYFAKEVSTTERTYL 556
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 597 RVVHSVQTFQEDSLRVAHSLEQVSEPLTAARHAHQKR-----635
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 557 KDLEV--ITSWFQSTVSKEDAMPEALKSLFFNPFELHFKHTNFKLEIQRALWEGRSN 614
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 636 NSIEIAGLTHGFEGLRKEETM-SASAAVAPRTFSAGSSQSVFSDKVQ-----EG--- 686
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 615 AQIRDYQIRIGDWMKNIQGMKHLAHLWKHSEALE---ALENGIKSSRLLENFCRDFELQ 671
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 687 ---TEEQSGSGGYSHKSLSDATMLIHSEDEDEDDSSREHAVSPRLTAAFSQQQLN 743
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 672 KYCYLPNTFLRLPLHRLMHYKQVLERLCKHHPSHADFRDCRAALAEITEMVAQLHGTM 731
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 744 YPC--ASVTPVTGPLH-----IFP-----PKSHVTEPEKRA-----KDISPVH--- 779
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 732 IKMENFQ-KLHELKDLIDNLVWPGREFIRLSLSK-----LSGK-----G 773
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 780 LVMETHQPRRHGLLTPSMSESDLTTSGRYARRRDSLLKRPVSDLLSGKKNTVEGLPLGG 839
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 774 LOORFEFL--FNDVLLYTSRGLTASNQFVKHQLPGLYGMTIESEDEWGVPHCLTLRGQR 831
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 840 MKKTRADAKKIGPLKLAALNGLSL-----RLPLPDEGKEVSTRATNDRCKVLEQRL 892
```

QY 832 OSIIVAASSRSEMEKVEDIQMAIDLAEKSSSPAPEFLASSPPDNKSPDEATRAQDSE 891
 Db 893 EQGTFTFYERILKRLVD-----GECSTARLPE-----NAERNFQD 930
 QY 892 DLSASRSLERQAPHRGNTVHVCHWRNTSVSMVDFS-IAVENOLSGNLLRKFNKNSGWQ 950
 Db 931 VLPYDDARVELVPTKENNTGYINASHIKVSVSGIEWDIATQGPLO-NTCDF-----WQ 984
 QY 951 KLVWFTNFCFLFFYKSHODNHPPLASLLGYSLTIPSESENIQDYVKLHFKHSHVYFR 1010
 Db 985 MVW-----EQGVAIAM-----VTAEEGGRE-----RSFRYWR 1014
 QY 1011 AEEPY---TFERNWEVIR-----SATSSASRPHVLSHKSLEY 1045
 Db 1015 LGSRHNTVTVYGRFKITTRFTDSCYATTGLKMKHLLTGQERTVM 1059

RESULT 15
 PTNL_MOUSE
 ID PTNL_MOUSE STANDARD; PRT; 1176 AA.
 AC Q62136;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE PTP-RL10).
 GN PTPN21.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=95140431; PubMed=7838537;
 RA Higashitsuji H., Arii S., Furutani M., Imamura M., Kaneko Y.,
 RA Takenawa J., Nakayama H., Fujita J.;
 RT "Enhanced expression of multiple protein tyrosine phosphatases in the
 RT regenerating mouse liver: isolation of PTP-RL10, a novel cytoplasmic-
 RT type phosphatase with sequence homology to cytoskeletal protein
 RT 4.1.1";
 RL Oncogene 10:407-414(1995).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF GROWTH AND
 CC DIFFERENTIATION OF LIVER CELLS.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + PHOSPHATE.
 CC -1- TISSUE SPECIFICITY: LIVER.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D37801; BAA07053.1; -
 CC HSSP: P18031; 1PTT.
 CC MGD: MGI:134406; Ptpn21.
 CC InterPro: IPR000299; Band_4.1.
 CC InterPro: IPR000387; Tyr_phosphatase.
 CC InterPro: IPR000242; Tyr_prot_phphatase.
 CC Pfam: PF00373; Band_41; 1.
 CC Pfam: PF00102; Y_phosphatase; 1.
 CC PRINTS: PR00700; PRTYPHPTASE.
 CC SMART: PR00935; BAND41.
 CC SMART: SM00295; B41; 1.
 CC SMART: SM00194; PTPc; 1.
 CC PROSITE: PS00660; BAND_41.1; 1.
 CC PROSITE: PS00661; BAND_41.2; 1.

DR PROSITE; PS50057; BAND_41.3; 1.
 DR PROSITE; PS50057; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 KW Structural protein; Cytoskeleton; Hydrolase.
 FT DOMAIN 77 241 BAND 4.1-LIKE
 FT DOMAIN 923 1176 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 1110 1110 BY SIMILARITY.
 FT DOMAIN 340 343 POLY-PRO.
 FT DOMAIN 565 572 POLY-PRO.
 SQ SEQUENCE 1176 AA; 133490 MW; 529FBE22F1335B75 CRC64;

Query Match 7.3%; Score 401.5; DB 1; Length 1176;
 Best Local Similarity 20.7%; Pred. No. 2.2e-15;
 Matches 248; Conservative 167; Mismatches 459; Indels 325; Gaps 45;

QY 25 LERGOKPPPTSGKLVSIKIOMLDDTQEAPEVPOAPGKVLDDAVCNHLNVEGDYGLE 84
 Db 9 LKTRRTYTVSKSCLVA-RQLLNNEVFETLSVESGQESLEAVORLEUREVYFSLW 67
 QY 85 FPDHKKITVMDLLKPIVKQIRRPKHVV---VKFVWAKFFPDHDTQLQBELTRYLFALQVK 141
 Db 68 YNKNQNRWVDLEKPLKKQL--DKHALEPTVYFGLVYFVPSVLSQLOQETRIYQYIQLK 125
 QY 142 QDLAQGRLTNDTSAALLISHIVOSEIGDFEALDRHLAKNKYIPOODALEKIVE--- 198
 Db 126 KDILEGNLPTCLHAIQLAGLAVQDFQDYESODFLQKFAALLPVAVLQDERVLEEA 185
 QY 199 ----FHHNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV-- 251
 Db 186 QKVALHQQYRGLTAPAEALYMOEVRMDGVEESTPAKDSQGSIDISIGACLDGIFVKH 245
 QY 252 FQGFTKINAFNAWRKLSFKRKLRIKRPDANSAYQDTLEFLMASRDFCKSKFWICVE 311
 Db 246 KNGRPV-VFRWHDIANMHNKSPFALEL---ANK--EETIQFOTEDMETAKYVRLCVA 299
 QY 312 HHAFRL-----FEEPKPK-----PKPVLFSRSGSFRFSRGT 344
 Db 300 RHKPYRLNOCSLQTAATLNSVRDSSRLPKPQPPQYAMPPPQLHYNGHYTPFPASSQ 359
 QY 345 KOVLDDYVKEG-GHKVKQFER-----KHSKIHSRS-----LASOP 378
 Db 360 DNIFVPKNKNGFYCHSQTSLDRTQIDLSGRINGSVYSAHSTNSLNTLQPYLPQSPMSNP 419
 QY 379 TELNSEVLE---SQSQSTSLTFEGEAESPGQS---CRRG----- 412
 Db 420 SITGSDVMRPDLSLPSHRHSALIPSYRPTPDYETVMKQLNKGWVHADRHSHSLRLNIGS 479
 QY 413 -----KEPKVSAGEPGSHP-----SPAPRRSPAGNKQADGAASA 446
 Db 480 SYAYSRPDALVYSQPEIREHPHLTSPQSAHYFPFNLTFSHSQSPYPYPAERPPVGVSV 539
 QY 447 PTEEEV-VKD-----RTQSKPPOPSTGLTSPHLS-ELSVNSGGVAPANVT 497
 Db 540 PELTNVQADQYAPNIMRTQVYRPPPPYPPRANSTPDLRHLIYSSN---PDLIT 596
 QY 498 LSPNLSPDTKQASPL-ISPLNDQACPRTDDEDEGRKRKRPDTPKAYFAKEVSTERTYL 556
 Db 597 RRVHHSVQTFQEDSLPVAHSLQEVSEPLTAARHAHLQKR-----NSIETAGLTHGFEGLRL 652
 QY 557 KLEVTTSWFQSTVSKEDAMPEALKSLIFPNFELHKEHFNFLKEIQRLALWGRSNAQ 616
 Db 653 KERTV-----SASAADVAPRTFSA-----GSSSV 677
 QY 617 IRDYQRIGDVMKNIQGMKHLAAHLWKHSALEALENGIKSRLENFCRDF----- 668
 Db 678 FSKMK-----QEGTEQEGGRYSHKKSLSDATMLDSEDEDEEDSSREQAISA 729
 QY 669 -----ELQKVCYLPLNTFLRLRHLRMHYKQVLERLCKHHPPSHADPRCRAA 716
 Db 730 VSEPLTAAFSQELNYPG--ASATPITGPHL-----IFE-----PKPHVTEPRKAK 774

```
QY 717 LAEITMVAQLHGTMIKMFNOKLHKKDLIGIDNLVVPGRFIRLGLSK-----LS 770
Db 775 DISPVHLVVETHR-----PRRDLTPSMSEDLTTSGRYRARRDSVKKRPVSDLLS 826
QY 771 GK-----GLQORMFEL--FNDVLLYTSRGLTASNQFKVHGQLPLYGWTIEESED 817
Db 827 GKKSAREGLPPLGGMKKTRADAKKIGPLKLAALNGLSL-----RLPLDDEGEVSTR 879
QY 818 ENGVPCHLFLRGOROSIIVAASSRSEMEKWEVQMAIDLAEKSSSPAPPEFLASSPPDNK 877
Db 880 ATNDERCKVLEORLEGGWFTYERILKRLVD-----GECSTARLPE----- 922
QY 878 SPDEATAADQESDDLSASRTSLERQAPHRGNTMVHVCHWRNTSYSMVDFS-IAVENQLS 936
Db 923 -----NAERNRFQDVL PYDDARVELVPTKENNTGVINASHIKVSVSGIEWDYIATQGPLQ 977
QY 937 GNLLRKFNSNGWQKLWVFTNFCLFFYKSHODNHPPLASPLLYSLTIPSESENIQKDY 996
Db 978 -NTCQDF-----WQWV-----EQVAIIAM-----VTAEEGGRE-- 1007
QY 997 VFKLHFKSHVYFRAESEY---TFERWMEVIR-----SATSSASRRPHVLSHKESLVY 1045
Db 1008 -----KSPRYWPRLGRSHNTVYGRFKITTRPTDSCGYATTGLKMKHLLTGQERTVW 1060
```

Search completed: December 6, 2001, 08:53:07
Job time: 379 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:49:58 ; Search time 77.11 Seconds
(without alignments)
1982.293 Million cell updates/sec

Title: US-09-555-342A-2
Perfect score: 5463
Sequence: 1 MGIEORPTGSLGAPENS.....SATSSASRPHVLSHKESLVY 1045

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.17.*
1: sp.archaea.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organelle.*
9: sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp.vertibrate.*
14: sp.unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5463	100.0	1045	Q9Y4F1	Q9Y4f1 homo sapien
2	2941.5	53.8	1054	Q94887	Q94887 homo sapien
3	2117	38.8	420	Q99KU3	Q99ku3 mus musculus
4	913.5	16.7	562	Q9VN69	Q9vn69 drosophila
5	867	15.9	330	Q9VN68	Q9vn68 drosophila
6	850.5	15.6	111	Q17905	O17905 caenorhabdi
7	746.5	13.7	929	Q9WV92	Q9wv92 mus musculus
8	746.5	13.7	963	Q9JMB2	Q9jmb2 rattus norv
9	743.5	13.6	865	Q9BRP5	Q9brp5 homo sapien
10	735.5	13.5	503	Q95713	O95713 homo sapien
11	732	13.4	988	Q170318	Q170318 mus musculus
12	728.5	13.3	1115	Q9Y2J2	Q9y2j2 homo sapien
13	728	13.3	1105	Q9JMB3	Q9jmb3 rattus norv
14	720	13.2	1005	O43491	O43491 homo sapien
15	711.5	13.0	706	Q9R102	Q9r102 mus musculus
16	708.5	13.0	864	Q9Y579	Q9y579 homo sapien
17	704	12.9	617	Q9N179	Q9n179 bos taurus
18	700	12.8	1551	Q9WVP1	Q9wvp1 rattus norv
19	699	12.8	879	Q9H4G0	Q9h4g0 homo sapien

20	698.5	12.8	641	4	Q14245	Q14245 homo sapien
21	695.5	12.7	879	11	Q9Z2H5	Q9z2h5 mus musculus
22	694	12.7	934	4	O15046	O15046 homo sapien
23	692.5	12.7	879	11	Q9WTP0	Q9wtp0 rattus norv
24	691.5	12.7	926	11	Q9WU22	Q9wu22 mus musculus
25	691.5	12.7	1698	5	Q24440	Q24440 drosophila
26	691.5	12.7	1698	5	Q9V8R9	Q9v8r9 drosophila
27	686.5	12.6	230	4	Q9UFE7	Q9ufe7 homo sapien
28	685	12.5	703	5	Q9V8S0	Q9v8s0 drosophila
29	682	12.5	889	5	Q9V8R8	Q9v8r8 drosophila
30	626	11.5	687	4	Q9H975	Q9h975 homo sapien
31	609.5	11.2	505	6	Q9MYU8	Q9myu8 canis famil
32	608	11.1	504	11	Q99KZ8	Q99kz8 mus musculus
33	586.5	10.7	972	5	Q9VFU8	Q9vfu8 drosophila
34	581	10.6	913	4	Q9H329	Q9h329 homo sapien
35	568	10.4	527	11	Q9JMC8	Q9jmc8 mus musculus
36	568	10.4	568	5	Q9GP93	Q9gp93 caenorhabdi
37	562	10.3	504	4	Q9H328	Q9h328 homo sapien
38	562	10.3	518	4	Q9P2V3	Q9p2v3 homo sapien
39	561	10.3	987	5	Q9NDP4	Q9ndp4 clona intes
40	534	9.8	440	4	Q9NX84	Q9nx84 homo sapien
41	524	9.6	283	4	Q9BQV2	Q9bqv2 homo sapien
42	520.5	9.5	766	11	O88387	O88387 rattus norv
43	518.5	9.5	598	4	Q9HCS5	Q9hcs5 homo sapien
44	465.5	8.5	414	5	Q20653	Q20653 caenorhabdi
45	460.5	8.4	699	4	Q9P215	Q9p215 homo sapien

ALIGNMENTS

RESULT 1

Q9Y4F1 PRELIMINARY; PRT; 1045 AA.
ID Q9Y4F1
AC Q9Y4F1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CDEP.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CARTILAGE;

RX MEDLINE=98086358; PubMed=9425278;

RA Koyano Y., Kawamoto T., Shen W., Yan W., Noshiro M., Fujii K.,

RA Kato Y.;

RT "Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors.";

RL Biochem. Biophys. Res. Commun. 241:369-375(1997).

DR EMBL; AB008430; BAA24267.1; .

DR HSSP; P08567; IPLUS.

DR InterPro; IPR000299; Band_4.1.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000219; RhoGEF.

DR Pfam; PF00373; Band_4.1; 1.

DR Pfam; PF00169; PH; 2.

DR Pfam; PF00621; RhoGEF; 1.

DR PRINTS; PR00935; BAND41.

DR SMART; SM00295; B41; 1.

DR SMART; SM00233; PH; 2.

DR SMART; SM00325; RhoGEF; 1.

DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.

DR PROSITE; PS50057; BAND_41_3; 1.

DR PROSITE; PS50003; PH_DOMAIN; 2.

SQ SEQUENCE 1045 AA; 118632 MW; 0E8B2D61C0F58417 CRC64;

Query Match 100.0%; Score 5463; DB 4; Length 1045;
Best Local Similarity 100.0%; Pred. No. 0;


```
Db 1 MGEIEGTYRVLOTAGMRLGAQTPGVSTLEPCQTLPRMQEKHLHLRVKLLDNTMEIFDI 60
Qy 57 QPARPGKVLDAVCHNLNLVGDYGLGEPDHHKITYWLDLLKPIVKQIRPKHVYVVF 116
Db 61 EPKCGQVLLTQVWRRLNVECDYGEFQNTQSWIWLPMKPIIQRIRPKVNLRLA 120
Qy 117 VKFPDPDHTQOEELTRYLFALQVQDLAQGLTNCNDSAAILLISHIVQSEIGDFEALD 176
Db 121 VKFPDPDQOQOEETRYLFALQRLDLLEERLTACADTAALLTSHLQSEIGDYDETL 180
Qy 177 REHLAKNYIPOQDALEDKIVFHHNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDRE 236
Db 181 REHLAKNYEYLPQOCHLEKILEFHQKHVGQTPAESDFQVLEIARLEMYGIRFHASDRE 240
Qy 237 GTKINLAVANTGILVFGQTKINAFNAWAKVKRLSKRFLIKLRDANSAYQDTLEFLM 296
Db 241 GTKIQLAVSHMGVLFQGTTKINTENWSKVKRLSKRFLIKLHPVHGVPYQDTLEFL 300
Qy 297 ASRDFCKFWKTCVEHHAFFRLEFEPKPKPVLFSRGSFRFSGRTOKQVLDYVKEGGH 356
Db 301 GSRDECKNFKVICVEYHTFFRLDQPKPAKAVFFSRGSSFRYSRGTOKQVLDYFKDSGM 360
Qy 357 KKVQPERKHSKIH-SIRSI-ASQPTLELSEVLEQSQSTSLTFGGAESPQSGQSCRR--- 411
Db 361 KRIPYERRHSKTHTSVRALTADLPQ-----SISFPEGLRTPASPSSANAFY 407
Qy 412 --GKEPKVSAGBPG-----SHPSPAPRRSPAGNKQADGAASAPTEEEVEVVKDRTOQ 461
Db 408 SLSPSTLVPGLPEFKDSSSLTDQVSVKSPAAERRSAGVAGGPD-----TPS 457
Qy 462 SKPQPP---QPTSGSLTGSPLSELVSNQGVAVANTVLSNLSPLDPTKQASPLISPLLN 518
Db 458 AQPLGPALQOPGLSTKSPQSPSSRKSPLSLSPA---FQVPLGPAEGSGSPLSPVL 514
Qy 519 DQACPRTDEDGRKRKPTDKAYIAKEVSTERTYLDLEIVTSWFQSTVSKEDAMPE 578
Db 515 DAGGAGMCE-EPRKRVPADEAYIIVEILATERTYLDLEIVTSFASVVKEDAMPA 573
Qy 579 ALKSLIFNFEPLKHFHTNFKLEIQRALWEGRSNAQIR-DYQRIGDVMKNIQMKHL 637
Db 574 TLMWTLFSNIDPIYEFHGFLEVEORLALWEGRSKAHTKGSHQIRIGDILLRNMQLKEF 633
Qy 638 AAHLWKHSEALEALENGIKSSRLENFCDFELQKVCYLPNTFLRLPHRLMHYKQVLE 697
Db 634 TSYFQRHDEVLELEKATKRCKKLEAVYKEFELQKVCYLPNTFLRLPHRLMHYKQVLE 693
Qy 698 RLCKHPPSHADFRDCRAALAEITWVAQLHCTMKMENFQKLHKLKDLIGDNLVVP 757
Db 694 RLCGHYSGHHDYACHDALKATEVTTLQHLIRLENLQKLTELQRLDVLGIENLIAPG 753
Qy 758 REFIRGLSKLSKGLQLOQRMFFLENDVLLYTSRGLTASNQKFKVHQPLPLYGMTIEESD 817
Db 754 REFIREGLHKLTKGLQLOQRMFFLENDVLLYTSKGVAGTSHFRIGLLPLQGLMVEESDN 813
Qy 818 EGVPHCLTLRGQROSTIIVAASSRSEMEKWKVEDIOMADLAEKSSSPAPEF---LASSP 873
Db 814 EWSVPHCFETIYAAQRTIIVAASTRLEKEKWLMDLSAIOAAKSGGDAPALPGRTVCTRP 873
Qy 874 PDKSPDEATAADOESEDLSASRTSLERQAPHRGNTMVHVCWHRNTVSMVDEFTAVEN 933
Db 874 P--RSPNE-VSUEQSEDDARVRSLEGGHQRANTTMHVCWYRNTSVSRADHSAAVEN 930
Qy 934 QLSGNLLRKFKNNSGWOKLWVYTNFCFFYKSHQDNHPLASLPLLGYSLTIPSESENIQ 993
Db 931 QLSGYLLRKFKNNSGWOKLWVYTNFCFFYKTHQDDYPLASLPLLGYSVSIPIREADGHI 990
Qy 994 KQYVFKLHFKSHVYFRAESETTFRMWEVIRSATSSASRPHVLSHESLVY 1033
Db 991 KQYVFKLQFKSHVYFRAESKYTFRMWEVIRSATSSASRPHVLSHESLVY 1030
```

RESULT 3
Q99KU3

```
ID Q99KU3 PRELIMINARY; PRT; 420 AA.
AC Q99KU3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3493093) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004009; AA004009.1; -.
FT NON_TER 1
SQ SEQUENCE 420 AA; 48387 MW; 804411A6B6613F96 CRC64;

Query Match 38.8%; Score 2117; DB 11; Length 420;
Best Local Similarity 95.2%; Pred. No. 1.4e-144;
Matches 400; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Qy 626 VMLKNIQGMKHLAAHLWKHSEALEALENGIKSSRLENFCDFELQKVCYLPNTFLRLP 685
Db 1 VMLKNIQGMKHLAAHLWKHSEALEALETSIGSRLEHFCDFELQKVCYLPNTFLRLP 60
Qy 686 LHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITWVAQLHCTMKMENFQKLHKLK 745
Db 61 LHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITWVAQLHCTMKMENFQKLHKLK 120
Qy 746 DLIGDNLVVPQREFIRGLSKLSKGLQLOQRMFFLENDVLLYTSRGLTASNQFKVHGL 805
Db 121 DLIGDNLVVPQREFIRGLSKLSKGLQLOQRMFFLENDVLLYTSRGLTASNQFKVHGL 180
Qy 806 PLYGMTIEESEDWGVPHCLTLRGQROSTIIVAASSRSEMEKWKVEDIOMADLAEKSSSPA 865
Db 181 PLYGMTIEESEDWGVPHCLTLRGQROSTIIVAASSRSEMEKWKVEDIOMADLAEKSNGET 240
Qy 866 PELLASSPDNKSPEATAADOESEDLSASRTSLERQAPHRGNTMVHVCWHRNTSVSMV 925
Db 241 PELLASSPDNKSPEATAADOESEDLSASRTSLERQAPHRGNTMVHVCWHRNTSVSMV 300
Qy 926 DFSIAVENOLSNLRKFNKNSGWOKLWVYTNFCFFYKSHQDNHPLASLPLLGYSLTI 985
Db 301 DFSIAVENOLSNLRKFNKNSGWOKLWVYTNFCFFYKSHQDNHPLASLPLLGYSLTI 360
Qy 986 PSESENIQDYVFKLHFKSHVYFRAESETTFRMWEVIRSATSSASRPHVLSHESLVY 1045
Db 361 PSESENIQDYVFKLHFKSHVYFRAESETTFRMWEVIRSATSSASRPHVLSHESLVY 420

RESULT 4
Q9VNE9 PRELIMINARY; PRT; 562 AA.
AC Q9VNE9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE CG2008 PROTEIN.
DE CG2008.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BERKELEY;
RL MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
```



```
QY 641 -LWKHSEALEALENGKSSRRLENFCRDFELQKVCLPLNTFLRLPHLRMHYKQVLERL 699
Db 760 VQAFHSETEVRE--IKTERASS-----HPLVHMOE--EIP 791
QY 700 CKHPPSHADFRCRAALAEITEWAQLHCTMIKMENFQKLHELKDLIDGLNVLVPGRE 759
Db 792 FSRAPLKFDKCKRSLSRALISVQSEDN----PDVQKCHLFSND-----IP--- 836
QY 760 FIRGLSLKSGKGLQORFFELFENDVLLVTSRG-----LTASNOFKVHGOLPL--- 807
Db 837 -----YILTMRVNVENTOSLPPSYSSFKDVSTAKKQGYESNLKRVSKSPFERRKSLDLVPRK 892
QY 808 -----YGMTIESEDEWGVPHCL-----TLRQROSIIVAASSRSEMEKWDIOM 853
Db 893 RLPSPGNFSAQDHTISPTTPDSVLEVLLRRRLGDK-SVIKTKSKRTDPRQTPVR- 950
QY 854 AIDIAKSSSPAPEFFLASPP-----DNKSPDEATAQDESDDLASRTSLERQAPH 906
Db 951 -FDLPSPSCSPT--AGGSTPFTISIINDDLFDECVSESRLHEDMD-----RLDKTTTPH 1000

RESULT 7
Q9WV92
ID Q9WV92 PRELIMINARY; PRT; 929 AA.
AC Q9WV92;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTEIN 4.1B.
GN EPB4.1L3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC TISSUE=BRAIN;
RA Parra M.K., Chan N.L., Gascard P.D., Walensky L.D., Lee G.L.,
RA Chasis J.A., Snyder S.H., Narla M., Conboy J.G.;
RT "Characterization of protein 4.1B, a new gene in the protein 4.1
RT family with high level, focal expression in brain.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152247; AAD38048.1; -
DR MGD; MGI:103008; Epb4.1l3.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
SQ SEQUENCE 929 AA; 10337 MW; F4975FF405DA44AE CRC64;

Query Match 13.7%; Score 746.5; DB 11; Length 929;
Best Local Similarity 32.6%; Pred. No. 3.3e-45;
Matches 205; Conservative 100; Mismatches 230; Indels 93; Gaps 22;

QY 18 ENSGISTLERGQ----KPPPTPSGKLVSKIOMLDDTQAEFVQAPGKVLVDVNCNHL 73
Db 97 QRSSSSKLSRSLPKIVKRP-----KSMQCKVTLLDGEYCDVDRKSRGQVLFDKVCEHL 151
QY 74 NLVEGDFYFGLFDPDHKKITVWLLDKLPVKQIRRPK-HVWVKFVWVFFPDHDTQLOEELT 132
Db 152 NLEKDYFGLTYDAENQKNWLDPAKEIKQIRSGAWH--FSFNVKFYPPDPAPQLSEDT 209
QY 133 RYLFALQVKODLAQGRLTCDNTSAAALLSHIVQSEIGDFD-EALDRHKLAKNKYIPQ-D 190
Db 210 RYLCQLQRRDVISGRPLPCSFVTVALLGSVTVQSELGVDYDPCDGCNDYISEFRFAPNHTK 269
QY 191 ALEDKIVFHHNIGQTPAESDFQLLEIARRLEMVGIHLPAKDRGTKLNLAVENTGIL 250
Db 270 ELEDKVIELHKSIRGMTPAEAMHFLLENAKKLSMYGVDLHAKDSGEVIMLGVCASGLL 329
```

```
QY 251 VFQGTFTINAFNWKVKLSFKRKRRELIKLRDPANSAYQDTLEFLMASRDFCKSFWKICV 310
Db 330 IYDRRLIRNFAMPKVLKISYKRNFFYIKIRPGCFQFQFESTIGFKLPNHRAAKRLMKVCV 389
QY 311 EHHAFPLRFEKPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKKVQFERKKSHIHS 370
Db 390 EHHTEFFRL-LPEAPPKFL-TLGSKFYSYSGRTQAQTRASALIDRPAPYFERSSSKRYT 447
QY 371 I-RSLASQTELNSEV-LEQSOOSTSLTGEAGSPGGQSC-----RRGKEPKVSA 419
Db 448 MGRSLDGA SVSENHETIYMKDSVAEVTGQYATTIGISQTNLITTVTEKKAEERVEE 507
QY 420 GE-----PGSHPSAPRR-----SPAGNKQADGAASAPTEEEVEVYK-----DRTQOS- 462
Db 508 EDRKKAEEATPVTLRHEGKTDSERTDAADGETSATESDQEDAEIKAAQDLDKTQDEL 567
QY 463 -KPQPPQPSGTSGTSPHLSSELSVNSQGGVAPAN-----VTLSPNLSPDTKQASPLISPL 516
Db 568 MKHQ-----TNISELKRFTLETSTALTNEWEKRLSTSPVRLAARQEDAPMIEPL 618
QY 517 LNDQACPRDDEDEGRR-----KRPETKAVFIAKEVSTTERTYTKDLEVIT 563
Db 619 V-----PEETKQSGGKLMGDSBILLESARKPTE---FIGGVSTTQ----- 659
QY 564 SWFQSTVSKEDAMPEALKSLIFNFEPL 591
Db 660 SWVQKLETKTPEVPEAEVEST--PHQPPL 685

RESULT 8
Q9JMB2
ID Q9JMB2 PRELIMINARY; PRT; 963 AA.
AC Q9JMB2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1 MINOR ISOFORM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC TISSUE=FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=20267856; PubMed=10806359;
RA Yamakawa H., Ohara O.;
RT "Comparison of mRNA and protein levels of four members of the protein
RT 4.1 family: The type II brain 4.1/4.1B/KIAA0987 is the most
RT predominant member of the protein 4.1 family in rat brain.";
RL Gene 248:137-145(2000).
DR EMBL; AB032828; BAA90775.1; -
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
SQ SEQUENCE 963 AA; 107071 MW; B37C137B04305F71 CRC64;
```

```
Query Match 13.7%; Score 746.5; DB 11; Length 963;
Best Local Similarity 33.7%; Pred. No. 3.5e-45;
Matches 198; Conservative 100; Mismatches 222; Indels 67; Gaps 18;

QY 18 ENSGISTLERGQ----KPPPTPSGKLVSKIOMLDDTQAEFVQAPGKVLVDVNCNHL 73
Db 92 QKSSSSKLSRSLPKIVKRP-----KSMQCKVTLLDGEYCDVDRKSRGQVLFDKVCEHL 146
QY 74 NLVEGDFYFGLFDPDHKKITVWLLDKLPVKQIRRPK-HVWVKFVWVFFPDHDTQLOEELT 132
Db 147 NLEKDYFGLTYDAENQKNWLDPAKEIKQIRSGAWH--FSFNVKFYPPDPAPQLSEDT 204
```

[illegible]

	Query Match	13.5%;	Score	735.5;	DB	4;	Length	503;
	Best Local Similarity	34.8%;	Pred.	No. 8.5e-45;				
	Matches	179;	Conservative	87;	Mismatches	209;	Indels	39; Gaps
								12;
Qy	43	KIOMLDTOEAFEPORAPGKVLDDAVCNHLNLVGEDYFGLFEFPDHHKITVWDLCLKPIV	102					
		: :: :	:	:	:	:	:	
Dd	4	KVILLDGSEYTCDEKRSGOVLFDKVCHEHNLLEKFGLTYRDAENQKNWLDPRAKEIK	63					
		: :	:	:	:	:	:	:
Qy	103	KQIRRPK-HVVVKVFVPFPDHTQLQEELTRYLFALOVKDLAGRLTCDNTSAAALLIS	161					
		:	:	:	:	:	:	
Dd	64	KQVRSGAWH--FSFNKVYPDPQAQLSEDITRYYLCLQRDDRIVSGRLPCSFVTLLALGS	121					
		: :	:	:	:	:	:	:
Qy	162	HIVOSEIGDFD-EALDRHLAKNKYIPQQ-DALEDKIVEFHNNHGICQTPAESDFOLLEIA	219					

```
Db 122 YTVQSEGLDYPDECGSDYISEFRFAPNHTKLEDEKVIELHKGSHRGMTPAEAEHMFLENA 181
QY 220 RRLMEYGLRHPAKDRGCTKINLAVANTGILVFQGTCKINAFNNAWKYKLFSEKRRKFLIK 279
Db 182 KKLISMYGVDLHHAADKSGEIMVGCASGLIYDRLRINRFAWPKVKLISYARNFYIK 241
QY 280 LRPDANSAYODTLEFLMASRDFCKSFWKICVEHHAFFRLEFEKPKPKVLFVSRGSSFRF 339
Db 242 IRPGEFFQFESTIGFKLPNRAAKRLWKVCVEHHTFRLL-LPEAPPKKFL-TLGSKFY 299
QY 340 SGRTOQVLYVKEGGHKVQFERKHSKINSI-RSLASQTELNSEV-LQSQOOSTSLTF 397
Db 300 SGRTOAQTRASALIDRPAPYFERSSSKRYTMSRSLDGASVNHETIYMDSMSAAEVGT 359
QY 398 GEAESPG-----GOSRRGCKPKVSAGEPGSHPSAPRSPAGNQAODG 442
Db 360 GOYATTKIGISQTNLITVTPEKKAEEERDEEDKRRKGEEVTPISAIQHEGKTDSEDT 419
QY 443 AASAPTEEEEVVKDRTOQSKPOPPQSTGLTSGPHLSVNSQGGVAPAN-----VT 497
Db 420 AADGETTATELEKTQDILMKHQ-----TNISELKRTFLETSTDTTAVTNEWKRLS 470
QY 498 LSPNLSPTQKASPLISPLNDQACRPTDDEG 531
Db 471 TSPVRLAARQEDAPMIEPLVPEEKM-ETKTESSG 503

RESULT 11
O70318 PRELIMINARY; PRT; 988 AA.
AC 070318;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN 4.1G (FRAGMENT).
GN EPB4.1L2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98198473; PubMed=9531554;
RA Walensky L.D., Gascard P., Fields M.E., Blackshaw S., Conway J.G.,
RA Mohandas N., Snyder S.H.;
RT "The 13-kD FK506 binding protein, FKBP13, interacts with a novel
RT homologue of the erythrocyte membrane cytoskeletal protein 4.1.";
RL J. Cell Biol. 141:143-153(1998).
DR EMBL; AF044312; AAC40083.1; -.
DR MGD; MGI:103009; Epb4.1L2.
DR InterPro; IPR000299; Band 4.1.
DR Pfam; PF00373; Band 4.1; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
FT NON_TER 988
SQ SEQUENCE 988 AA; 109833 MW; B0367A16C5A2EC05 CRC64;

Query Match 13.4%; Score 732; DB 11; Length 988;
Best Local Similarity 32.7%; Pred. No. 4e-44;
Matches 198; Conservative 91; Mismatches 229; Indels 88; Gaps 18;

QY 34 TPSEGLVSTKIOMLDDTQAEFVQAPQKVLVDVNCNHLNLVEGYFGLFPDHPKHTIV 93
Db 205 TKTKTKVLAKVTLTDGTEYSCDLKRAKQVLFDRVCEHLNLEKDYFGLLFQDHPQKN 264
QY 94 WLDLLPIYVKVIRRPKHVVVVFVFPDPDHTQLOQLRYLQVLAQGRGLTCND 153
Db 265 WLDPAKEIKRQLKNLPLWFT-FNVKYPDPDSOLTDITRYFLCLQLRODIASGRLPSCF 323
```

```
QY 154 TSAALLISHIVQSEIGDFD-EALDREHLAKNKYIP-QQDALEDKIVEFHNNHIGOTPAES 211
Db 324 VTHALLSGSYTLQAEHGDYDPEEYDSIDLDFQFAPAHKTELEEKSELHKHTRGLSPAQA 383
QY 212 DFOLLEIARLEMYGRLHPAKDRGCTKINLAVANTGILVFQGTCKINAFNNAWKYKLF 271
Db 384 DSQFLENAKRLSMYGVLDLHHAADKSGEIMVGCASGLIYDRLRINRFAWPKVKLISY 443
QY 272 KRKRELILRPDANSAYODTLEFLMASRDFCKSFWKICVEHHAFFRLEFEKPKPKVLF 331
Db 444 KRNFYIKVRPAELEQFESTIGFKLPNRAAKRLWKVCVEHHTFYRLV-SPEQPPKTKFL 502
QY 332 SRGSSFRFGSRTQKQVLDVYKGGCHKVQPERKHSKINSI-RSLASQTELNSEVLEQSOQ 391
Db 503 TLGSKFYRSRTQAOQTRASTLIDRPAPQFERASSKRV-RSLDGAPI-----GVQDSPP 557
QY 392 STSLTTEGA-ESPGQSGRRGKEPKVSAGEPGSHPSAPRSPAGNQAODG-----442
Db 558 -----GEGSVPGPGVISYTTIQDGRDSSKSP-TKATPLPAEGKKNLTVRVDGNIYVRS 610
QY 443 -----AASAPTEEEEVVKDRTOQSKPOPPQSTGLS 474
Db 611 NLMLDLDKAQEAAILKHQASISELKNFNASTPEPRPSEWEKRRV---TLPPLQPAQS- 666
QY 475 TGSPHLSVNSQGGVA--PANVT-----LSPNLSP-DTKQASPLI---SPLNDQA 521
Db 667 --HETLVNVEEKKRAGVKGDESIVTEEMNGKEMSPGPGETRKVEPVAHKDSTLSSES 724
QY 522 CPRTDDEGRRKRPDTPDKAYFAKEVSTTERTYTKDLEVITSWFQ-----STVSKED 574
Db 725 SSSSESEEDVGEYQPHHR-----VTEGTIREQECEDELEPEEPQGGAKVVEREA 775
QY 575 AMPEAL 580
Db 776 AVPDV 781

RESULT 12
QY2J2 PRELIMINARY; PRT; 1115 AA.
ID QY2J2;
AC QY2J2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIA0987 PROTEIN (FRAGMENT).
GN KIA0987.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023204; BAA76831.1; -.
DR InterPro; IPR00299; Band 4.1.
DR Pfam; PF00373; Band 4.1; 1.
DR PRINTS; PR00935; BAND41.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
DR SMART; SM00295; B41; 1.
FT NON_TER 1115
SQ SEQUENCE 1115 AA; 123704 MW; E38E3A31E028D4A2 CRC64;
```

```
Query Match      13.3%; Score 728.5; DB 4; Length 1115;
Best Local Similarity 32.1%; Pred. No. 8.6e-44;
Matches 196; Conservative 92; Mismatches 217; Indels 105; Gaps 19;

QY 18 ENSGISTLERG-----OKPPPTSGKLVSIKIQMLDDTQEAEPVQAPGKVLDDAVCNH 72
DB 117 QSSSSKLSRSLKIVKRP-----KSMCKVILLDGSYTCDEKRSQGVLFQKVCBH 170

QY 73 LNLVEGDYFLEFPDHHKTIIVWLDLPIVKQIRRPK-HVVVVFVVFVFPDHTQLEEL 131
DB 171 LNLLEKDYFGLTYRAENQKNWLDPAKEIKQVRSAGWH--FSFNVKFPYPPDPAQLSEDI 228

QY 132 TRYLFALQVKQDLAOGRLTCDNTSAALLISHIVQSEIGDFD-EALDREHLAKNKYIPQO- 189
DB 229 TRYILCLQLRDDIVSGRLPCSFVTLLALLGSYTVQSELGDYDDEGSDYISEFRFAPNHT 288

QY 190 DALEKIVFEHHNHIGOTPAESDFOLLETARLEMYGIRLHPAKDREGTKINLAVANTGI 249
DB 289 KELEKVIELHSHRGMTPAEAEHMFLENAKKLSMYGVDLHAKDSEGVIMLGVCASGL 348

QY 250 LVFOGFTKINAFNNAKVRKLSFKRRLIKLRDPDANSAYQDTLEFLMASRDFCKSPWKIC 309
DB 349 LIYRDLRLINRFAMPKVLKISYKRNFIKIRPGEFEQFESTIGFKLPNHRAAKRLWKVC 408

QY 310 VEHAFRLFEPEPKPKPVLSRSRGSFSGRTQKVLDYVKEGGHKVQFERKHSKTH 369
DB 409 VEHTHTFRLL-LPEAPPKKFL-TLGSKFYSRGTQAOTRRASALIDRPAPYFERSSSKRY 466

QY 370 ST-RSLASQPTELNSVLRSQOSTSLTGTGEAESPGGSCRGKPKYSAGEPGSHSP 428
DB 467 TMSRSL-----DGEVGTGQYATTKG-----ISQTNLIITV 496

QY 429 APRRSAGNKQADGAASAPTEEEVVKDRTOQSKPQPPQSTGSLTSPHLSSELSVNSQ 488
DB 497 TPEKK-----AEERDEEDKRRKGEVTP-----ISAIRH 527

QY 489 GGVPANVTLSPLNLPDTPKQASPLISPLNDQACPRTDDE----DEGRKRPRPTDKAYFI 544
DB 528 EGKSPGLGDSCLPLSPPTSHCAP-TSPTELRRCKENDCKLPGYEPSRAEHLPGEPAL-- 584

QY 545 AKEVSTTERTYTKDLEVITSWFQST-----YSKEDAMPEALKSLI-----FPNFEPL 591
DB 585 --DSGDGPRPYLGDQDVAFSYRQAGKGTTLFSLSQLPESFPLDDGYSFPNLS-- 640

QY 592 HKFHTNFKLE 601
DB 641 ---ETNLLPQ 647

RESULT 13
QJUMB3
ID QJUMB3 PRELIMINARY; PRT; 1105 AA.
AC QJUMB3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RX MEDLINE=20267856; PubMed=10806359;
RA Yamakawa H., Ohara O.;
RT "Comparison of mRNA and protein levels of four members of the protein
RT 4.1 family: The type II brain 4.1/4.1B/KIAA0987 is the most
RT predominant member of the protein 4.1 family in rat brain.";
RL Gene 248:137-145(2000).
DR EMBL; AB032827; BAA90774.1; -
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
```

```
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00507; BAND_41_3; 1.
SQ SEQUENCE 1105 AA; 122451 MW; 94F1CA93A83693C6 CRC64;

Query Match      13.3%; Score 728; DB 11; Length 1105;
Best Local Similarity 32.6%; Pred. No. 9.2e-44;
Matches 198; Conservative 97; Mismatches 225; Indels 88; Gaps 20;

QY 18 ENSGISTLERGQ-----KPPPTSGKLVSIKIQMLDDTQEAEPVQAPGKVLDDAVCNH 73
DB 92 QSSSSKLSRSLKIVKRP-----KSMCKVILLDGSYTCDEKRSQGVLFQKVCBH 146

QY 74 LNLVEGDYFLEFPDHHKTIIVWLDLPIVKQIRRPK-HVVVVFVVFVFPDHTQLEELT 132
DB 147 LNLLEKDYFGLTYRAENQKNWLDPAKEIKQIRSGAWH--FSFNVKFPYPPDPAQLSEDI 204

QY 133 RYLFALQVKQDLAOGRLTCDNTSAALLISHIVQSEIGDFD-EALDREHLAKNKYIPQO-D 190
DB 205 RYILCLQLRDDIVSGRLPCSFVTLLALLGSYTVQSELGDYDDEGSDYISEFRFAPNHTK 264

QY 191 ALEDKIVFEHHNHIGOTPAESDFOLLETARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 265 ELEDKVIELHSHRGMTPAEAEHMFLENAKKLSMYGVDLHAKDSEGVIMLGVCASGL 324

QY 251 VFOGFTKINAFNNAKVRKLSFKRRLIKLRDPDANSAYQDTLEFLMASRDFCKSPWKICV 310
DB 325 IYRDLRLINRFAMPKVLKISYKRNFIKIRPGEFEQFESTIGFKLPNHRAAKRLWKVCV 384

QY 311 EHHAFRLFEPEPKPKPVLSRSRGSFSGRTQKVLDYVKEGGHKVQFERKHSKTHS 370
DB 385 EHTHTFRLL-LPEAPPKKFL-TLGSKFYSRGTQAOTRRASALIDRPAPYFERSSSKRYT 442

QY 371 I-RSLASQPTELNSEV-LEQSOOSTSLTGTGEAESPGGSCRGKPKYSAGEPGSHSP 428
DB 443 MRSLDGASVNHENHETLYMKDSMAAEVGTGQYATT-----KVIS----- 481

QY 429 APRRSAGNKQADGAASAPTEE--EEVVKDRTOQSKPQPPQSTGSLTSPHLSSELSVN 486
DB 482 -----QTNLIITVTPKAEHEEVEDRRKKAETP-----VAAL 518

QY 487 SOGGVAPANVTLSPLNLPDTPKQASPLISPLNDQACPRTDDEDEGRKRPRPTDKAYFI 546
DB 519 RHEGKSPGHGSDCLPLSPPTSHALDP-PSPTELRRCKEKERASPG--SAVPRAE--LSS 572

QY 547 EVSTTERTYTKDLEVITSWFQ-----SVSKEDAMPEALKSLI-----FPNFEPLHK 593
DB 573 ESDAQGKAYLGDQDVAFSYRQAGKGTTLFSLSQLPESFPLDDGYSFPNLS---- 628

QY 594 FHTNFKLE 601
DB 629 -ETNLLPQ 635

RESULT 14
O43491
ID O43491 PRELIMINARY; PRT; 1005 AA.
AC O43491;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN 4.1-G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98260860; PubMed=9598318;
RA Parra M., Gascard P., Walensky L.D., Snyder S.H., Mohandas N.,
```

RA Conboy J.G.;
RT "Cloning and characterization of 4.1G (EPB41L2), a new member of the
RT skeletal protein 4.1 (EPB41) gene family.";
RL Genomics 49:298-306(1998).
DR EMBL: AF027299; AAC16923.1; -.
DR InterPro: IPR000299; Band_4.1.
DR Pfam: PF00373; Band_41; 1.
DR PRINTS: PR00935; BAND41.
DR PROSITE: PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE: PS00661; BAND_41_2; 1.
DR PROSITE: PS00657; BAND_41_3; 1.
DR SMART: SM00295; B41; 1.
SQ SEQUENCE 1005 AA; 112587 MW; E86CB17488F6045F CRC64;

Query Match 13.2%; Score 720; DB 4; Length 1005;
Best Local Similarity 31.2%; Pred. No. 3e-43;
Matches 192; Conservative 102; Mismatches 222; Indels 100; Gaps 18;

QY 34 TPGSKLVSTKIOMLDDTQAEFVQAPGKVLDDAVCNHLNVEGDYFGLFEPDHHKTIIV 93
Db 212 TKTKTVQCKVLLDGTETSCDEKHAQGVLFDKVCEHLNLEKDYFGLLFQESPEQKN 271
QY 94 WDLLEPIVQIRRPKHVVVVFVFFPPDHTQLQBELTRYLFALQVQDLAQRILCND 153
Db 272 WLDPAKEIKRQLRNLPWLT-FNVKFPDPSPQLTEDITRYFLCLQRLQDIASGRLEPCSF 330
QY 154 TSAALLSHIVQSEIGDFEALDREH---LAKNKYIP-QQDALEDKIVEFFHHHLCQTP 208
Db 331 VTHALLGSTYQALQELDYD---PEEHSIDLSFQAPDTQTELEKVAELHKTHTGLSP 387
QY 209 AESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNNAKVRK 268
Db 388 AQADSQFLENAKRLSYGVDLHAKDSEGVQIKLVCANGLLIYKDRILNRFPANFKILK 447
QY 269 LSPKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRLEFPKPKPK 328
Db 448 ISYKRSNFYIKVPAELEQFESTIGFKLPNHRAAKRLKVCVEHHTFYRLV-SPEQPPKA 506
QY 329 VLFSGSSRFSGRTQKQVLDYVKEGHHKVKQPERKHSKIHRSLSASQTELNSEVLEQ 388
Db 507 KFTLGSKFRYSRTQAQTRQATSLDIRPAPHFRTSSKRVLSRLDGAPI---GVMDQ 561
QY 389 SQGSTSLTFCGEAESPGGSCRRGKPKVSAGPSGSHPSAPRRSPAGNKQADGAASAP- 447
Db 562 SLMK-----DFPGAAG-----EISAYGPGI-VSIAYVQDGDGRREVRSPTKAPH 604
QY 448 -----TEEEEVVKDRT-----T-----QOSKPPPP- 467
Db 605 LQIEGKNSLRVEGDNIVYRHSNLMLELDKAQEDILKHOASISELKRNFMESTPEPRP 664
QY 468 -----OPSTGSLTGSPPH---LSLSVNSOGGVAPANVT-----LSPNLSP-DTK 507
Db 665 NEWEKRRTIPLSLQTCGSHSETINIVEEKKRAEVKDERVITEEMNGKELSGSGGPEIR 724
QY 508 QASPLI-----SPLINDQACPRTDEDEGRKRKRPPTKAYFIAKEVSTTERTYILKOLEVIT 563
Db 725 KVEPVTKDSTSLSSSSSSSESEEDVGEYRPHR---VIEGTIRREQVEEVEEPEP 781
QY 564 SWFQSTVSKEDAMPEA 579
Db 782 RPAKVVEREEAVPEA 797

RESULT 15
QYR102 PRELIMINARY; PRT: 706 AA.
ID QYR102
AC QYR102;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE DALI.
GN DALI.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRAIN;
RA Azam M., Andradi S., Lin L., Newsham I., Chishti A.H.;
RT "Mouse DAL-1 (MDAL-1) cDNA Sequence."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF177146; AAD51365.1; -.
DR InterPro: IPR000299; Band_4.1.
DR Pfam: PF00373; Band_41; 1.
DR PRINTS: PR00935; BAND41.
DR PROSITE: PS00660; BAND_41_1; 1.
DR PROSITE: PS00661; BAND_41_2; UNKNOWN_1.
DR PROSITE: PS00657; BAND_41_3; 1.
DR SMART: SM00295; B41; 1.
SQ SEQUENCE 706 AA; 79448 MW; 75B6A3EF475A5E31 CRC64;

Query Match 13.0%; Score 711.5; DB 11; Length 706;
Best Local Similarity 32.4%; Pred. No. 7.5e-43;
Matches 185; Conservative 99; Mismatches 250; Indels 37; Gaps 12;

QY 43 KIQMLDDTQAEFVQAPGKVLDDAVCNHLNVEGDYFGLFEPDHHKTIIVWLDLKP 102
Db 4 KVTLLDSEYGVCDVRSRGQVLFDKVCEHLNLEKDYFGLTYRDAENKKNWLDPAKEIK 63
QY 103 KQIRRPK-HVVYKVVVVFVFFPPDHTQLQBELTRYLFALQVQDLAQRILCNDTSAALLIS 161
Db 64 KQIRSGAWH-FSFNVKFPDPQALSEDITRYFLCLQRLDDIVSGRLPCSFVTALLGS 121
QY 162 HIVQSEIGDFD-EALDREHLAKNKYIPQ-QDALEDKIVEFFHHHLCQTPAESDFOLLETA 219
Db 122 YTVQSEIGDYDEDCNDYISEFRFAPNHTKELEDKVIELHKSHRGMTPEAEHMFLENA 181
QY 220 RLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNNAKVRKLSFKRRFLIK 279
Db 182 KKLMSYGGDLHAKDSEGVQIKLVCANGLLIYRDLRINRFPANPKVLKISYKRNFIYK 241
QY 280 LRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRLEFPKPKPKPVLFSGSGSFRF 339
Db 242 IRPGEPEQFESTIGFKLPNHRAAKRLKVCVEHHTFRL-LPEAPPKKFL-TLGSKFY 299
QY 340 SGRTOQVLDYVKEGHHKVKQPERKHSKIHSL-RSLASQ---PTELNSEVLEQSQSTSL 395
Db 300 SGRTOQVTRASALIDRPAPYFERSSSKRYTMSRLDGEVGTGQYATTKGISQTNLITV 359
QY 396 TFCGEAESPGGSCRRGKPKVSAGPSGSHPSAPRRSPAGNKQADGAASAPTEEEVEV 455
Db 360 TPEKKAEEB-----RVEEDRRKKAEEATPVIALRHEGKTDSERTDAADGETSATEDLD 414
QY 456 KRTQOSKQPPQPPQSTGSLTSPHLSLSVNSOGGVAPAN-----VTLSPNLSPTDKQAS 510
Db 415 KQDELMLKHQ-----TNISELKRTELETSTALTNEWKRLSTSPVRLAARQEDA 465
QY 511 PLISPLINDQACPRTDEDEGRKRKRPPTKAYFIAKEVSTTERTYILKOLEVITSWFQSTV 570
Db 466 PMIEPLVPEEKLETKTEPVEAEVESTPHPOPL-----STEKVLOETILVREHVMVSH 518
QY 571 SKEDAMPEALKSLIFFNFPFELHFKFTNFKLE 601
Db 519 ASGDASHARTARDEVDAAESTPTDRRHTGKGE 549

Search completed: December 6, 2001, 08:54:30
Job time: 272 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 12:06:44 ; Search time 4659.49 Seconds
(without alignments)
11099.639 Million cell updates/sec

Title: US-09-555-342A-1_COPY_49_3183

Perfect score: 3135

Sequence: 1 atggggaataagacagag.....acaaagagtctctgtgtat 3135

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vl.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	3135	100.0	3442	9	AB008430	AB008430 Homo sapi
2	1043	33.3	2632	10	BC004009	BC004009 Mus muscu
3	1037.4	33.1	3997	9	AB018336	AB018336 Homo sapi
4	1013.2	32.3	3719	10	BC009153	BC009153 Mus muscu
5	342.8	10.9	2888	9	HSN802608	AL161984 Homo sapi
6	286.6	9.1	1438	9	HSN801276	AL122052 Homo sapi
7	254	8.1	96183	9	AL161896	AL161896 Human DNA
c	254	8.1	152053	2	AC022669	AC022669 Homo sapi
8	254	8.1	181800	2	AL137249	AL137249 Homo sapi
9	254	8.1	181800	2	AL137249	G22662 human STS W
c	210	6.7	483	11	G22662	AF156225 Homo sapi
10	194	6.2	2595	9	AF156225	AC022669 Homo sapi
11	193.4	6.2	152053	2	AC022669	M14993 Human struc
12	193.4	6.1	2867	9	HUMELIA	M61733 Homo sapien
13	192.4	6.1	3043	9	HUMEMP41	L00919 Mus muscullu
14	192.4	6.1	3043	9	HUMEMP41	AF106702 Mus muscu
15	189.4	6.0	5942	10	MUS41SP	AB019256 Rattus no
16	187	6.0	2882	10	AF106702	AB019257 Rattus no
17	185.2	5.9	6102	10	AB019256	BC010674 Homo sapi
18	183.6	5.9	5855	10	AB019257	M68941 Human prote
19	181.8	5.8	3619	9	BC010674	AF222767 Bos tauru
20	181.8	5.8	3643	9	HUMPTYPH	AB032828 Rattus no
21	181.2	5.8	2590	4	AF222767	AB032827 Rattus no
22	181.2	5.8	3661	10	AB032828	L27467 Drosophila
23	181.2	5.8	4543	10	AB032827	AF044312 Mus muscu
24	180.4	5.8	6263	9	AB002336	AF061283 Mus muscu
25	179.6	5.7	5921	3	DROP41A	AL336300 Human DNA
26	176.6	5.6	2964	10	AF044312	M20621 X.laervis cy
27	174	5.6	3320	10	AF061283	AF027299 Homo sapi
28	172.8	5.5	170916	9	ALI36300	AB032366 Mus muscu
29	172.4	5.5	2758	5	XELCSK	AF152247 Mus muscu
30	169.6	5.4	4336	9	AF027299	G60059 SHGC-130885
31	166.8	5.3	3451	10	AB032366	AF177146 Mus muscu
32	165	5.3	4051	10	AF152247	AF069072 Homo sapi
c	162.4	5.2	393	11	G60059	BC006141 Homo sapi
33	162.4	5.2	393	11	G60059	AB023204 Homo sapi
34	161.8	5.2	2717	10	AF177146	AB070172 Macaca fa
35	161.6	5.2	3309	9	AF069072	J03796 Human eryth
36	161.6	5.2	3387	9	BC006141	AC018327 Drosophil
37	161.6	5.2	4446	9	AB023204	AC009537 Drosophil
38	157.6	5.0	1973	9	AB070172	AC007589 Drosophil
39	154.2	4.9	2677	9	HUMELI	AE003604 Drosophil
40	154	4.9	42014	2	AC018327	M64572 Human prote
41	154	4.9	160710	3	AC009537	BC011476 Mus muscu
42	154	4.9	164713	3	AC007589	
43	154	4.9	303209	3	AE003604	
44	144	4.6	3984	9	HUMCAP	
45	143.2	4.6	2211	10	BC011476	

ALIGNMENTS

RESULT 1
AB008430 LOCUS AB008430 3442 bp mRNA PRI 13-FEB-1999
DEFINITION Homo sapiens mRNA for CDEP, complete cds.
ACCESSION AB008430
VERSION AB008430.1 GI:2766164
KEYWORDS CDEP.
SOURCE Homo sapiens embryo cartilage chondrocyte cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3442)
AUTHORS Koyano,Y., Kawamoto,T. and Kato,Y.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-1997) to the DDBJ/EMBL/GenBank databases. Takeshi
Kawamoto, Hiroshima University School of Dentistry, Department of
Biochemistry; 1-2-3 Kasumi Minami-ku, Hiroshima, Hiroshima 734,
Japan (E-mail:tkawamo@ipc.hiroshima-u.ac.jp, Tel:082-257-5688,
Fax:082-257-5629)
2 (sites)

AUTHORS Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and Kato,Y.
 TITLE Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of rho guanine nucleotide exchange factors
 JOURNAL Biochem. Biophys. Res. Commun. 241 (2), 369-375 (1997)
 MEDLINE 98086358
 FEATURES
 Location/Qualifiers
 1..3442
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="chondrocyte"
 /dev_stage="embryo"
 /tissue_type="cartilage"
 49..3186
 /function="Rho Guanine Nucleotide Exchange Factor"
 /note="Band 4.1 superfamily"
 /codon_start=1
 /product="CDEP"
 /protein_id="BAA24267.1"
 /db_xref="GI:2766165"
 /translation="MGEIEQRTPGSRIGAPENSGISTLERQKPPPTSPGKLVSIKI QMDDTFAFVQPARPGKVLVDVNCNHLNLVFGDYFGLFDPDHKKITVWLDLLKPIV KQIRPKHVVVVFVFPDPHTQLQELTRYLFALQKDLAQRLTNCNDTSAALLI SHVQSEIGDFDEALDRHLAKNKV IPOODALEDKIVEFHNNHIGOTPAESDFOLLEI ARLEWYIGRLHPADRECTKINLAVANTGILVFOGFTKINAFNAKVKFLSKRKR LKLRPDANSAYODTLEFLMASRDFCKSPWKICVSEHAFPRFLPEPKPKPVLSRG SFRSGRTQKQVLDIVRGKHKVQFERKHSKIHISLSQSPTELVSEVLEQSQS TSLTGEAGSEGGSCRRKEPKVSAGEPGSPAPRRSPAGNKAQDGAASPTKEE EVVVKDRTQKQPPQPSGTSLGSPHLSVNSQSGVAPANTVLSPLSDPTKQ ASPLSLINDQACPTDDEDEGRKRFPDKAYTAKEVSTERTYLLKDLVITSWF QSTVSKEDAMPALKSLIFPNPEPLKHTNFKLEORLALWEGRSNAQIRDYORIG DYMKNIOGMKHLAHLKHSALALENGIKSRRLNFCRDFELQKCYLPPLTFLL LRLRLHMKVLERLCKHPPSHADFDRCRAALAEITEMVAUHGTMKNENFQKL HEKDKLIGIDNLVVGREFIRGLSKSLGSLGQLORMFFLFDNVLVLTSLRGLTASQ FVKHQGLPLYGNTIESEDEWGPVHCLTLRGQRSIIIVAASRSRMEKWEVDIOMAD LAEKSSPAPELASPPDNKSPDEATAADQSEDDLSASRTSLERQAPHRGNTWVHV CWHRTSVSMVDFSLAVENQLSKNLLRKFKNWGKLVWVETNCLFFPKSHOHNHP LASPLGLSLTIPSESENIQDVFELFKHSHVYFFRAESYTERWMEVIRSATSS ASRPVLSHKSLSVY"
 polyA_site 3442
 /note="50 a nucleotides"
 BASE COUNT 864 a 952 c 927 g 699 t
 ORIGIN
 Query Match 100.0%; Score 3135; DB 9; Length 3442;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgggagaaatagacagagcgacccagatcacgactcagggcgcccgcaaaattcgg 60
 Db 49 ATGGGAGAAATAGACAGAGCGCCAGCCAGGATCAGACTGGGGCGCCGAAAAATCG 108
 QY 61 gggatcagttaccttggaaacgtggacagaagcgcccccacaccttcaggaaaaactcgtg 120
 Db 109 GGGATCAGTACCTTGGAACTGGACAGAGAGCGCCGCCCAACACCTTCAGGAAACTCGTG 168
 QY 121 tccatcaaatccagatgctgtagtgatcacacccagagggcatttgaagttccacaaagagct 180
 Db 169 TCCATCAAAATCCAGATGCTGGATGATACCCAGAGGCGCATTTGAAGTTCACAAAGAGCT 228
 QY 181 cctgggaagtgctgctgtagtcagttgtcaacacccactcaacctcgtgaagtgactat 240
 Db 229 CTTGGGAAGTGCTGCTGATGATGAGTTGTCAACACCTCAACCTCGTGGAGGTGACTAT 288
 QY 241 ttggcctcagtttctctgatcacaaaaagatcacgggtggtggtgtagtctctcaaaacc 300
 Db 289 TTTGGCCTCGAGTTCTCTGATCACAAAAAGATCACGGTGTGGCTGGATCTCTTAAACCC 348
 QY 301 attgtgaacagattagaaggccaaagcagctgtgtgttaagtgtgtgtaattcttt 360
 Db 1429 CAGAGTAAACCTCAGCCCCCGCCGACCAAGCACAGGCTCCCTTGACTGGCAGTCTCACCCTT 1488

Db 349 ATTGTGAAACAGATTAGAAGGCCAAAGCACGTTGTTGTTAAGTTTGTGTGTAATCTTTT 408
 QY 361 cgcctgacacacacacacactccaagaagaactcacaaaggtacactgttcgcgtgcagtg 420
 Db 409 CCCCTTGACCACACACAACTCCCAAGAAGAACTCACAAAGGTACTGTTCCGCGCTGCAGGTG 468
 QY 421 aagcaggaacttggtcctcaagcaggttgacgtgtaataagacaccagcgcagctctcttgatt 480
 Db 469 AAGCAGGACTTGGCTCAAGCGCAGGTTGACGTGTAATGACACACGCGCAGCTCTCTTGATT 528
 QY 481 tccacattgtgcaatctcagatttggggatttggatgaagccttgacagagagcactta 540
 Db 529 TCACACATTGTGCAATCTCAGATTGGGGATTTTGTGTAAGCCTTGGACAGAGACACTTA 588
 QY 541 gcaaaaaataataacataactcagcaagcagcactagagagcaaaaatcgtggaaattcac 600
 Db 589 GCAAAAAATAATATACATACCTCAGCAAGCGCACTAGAGGACAAATCTGTTGGAAATTTTCA 648
 QY 601 cataaccacattggacaacacaccagcagatcagatttccagctccttagagattgcccgt 660
 Db 649 CATAACCACATTGGACAACACACGAGCAATCAGATTTCAGCTCTCTAGAGATTGCCCGT 708
 QY 661 cggctagatgtatggaatccoggttgccccggcccaagagacagggagagcagaagatc 720
 Db 709 CGGCTAGAGATGTATGGAATCCGTTGCACCCGCCCAAGACAGGGAAGCACGAAGATC 768
 QY 721 aactcgccgttgcccaacacgggaattctagttttcaggggtttcactaagatcaatgcc 780
 Db 769 AATCTGGCCGTTGCCAACACGGGAATTTCTAGTGTTCAGGGTTTCACTAAGATCAATGCC 828
 QY 781 ttcaactgggccaaggtgcggaagctgagcttcaagaggaagcgcttctcatcaagctc 840
 Db 829 TTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTC 888
 QY 841 cggccagatgcgaatagtcgtaccaggatcaccttggaattctcgtatggtgcagtcggat 900
 Db 889 CGGCCAGATGCCAATAGTGCCTACCAAGATACCTTGGAAATTCCTGTATGGCCAGTCGGAT 948
 QY 901 tctcgaagctctctggaaaatctgtgtgaacatcagctcctcttttagactttttaa 960
 Db 949 TTTCTGAAGTCTCTTGGAAAAATCTGTGTTGAACATCATGCTCTTTAGACTTTTGTAA 1008
 QY 961 gagcccaaaccaagccccagccgctcctcttttagccgggggttcattcatttccggttcagt 1020
 Db 1009 GAGCCCAACCAAGCCCAAGCCCGTCTCTTTAGCCGGGGTCACTCATTTCCGTTCACT 1068
 QY 1021 ggtcgactcagaagcaggttctcagactatgttaagaagagagagacataagaaggctcag 1080
 Db 1069 GGTGGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAGGAGGACATAAGAAGGTGCAG 1128
 QY 1081 ttgaaaggaagcacacagcagattcattctccggagccttgcttcacagcctacagaa 1140
 Db 1129 TTTGAAAGGAGCACACAGCAAGATTCAFTCTATCCGAGAGCTTGTCTTCACAGCTACAGAA 1188
 QY 1141 ctgaattcgggaagtcgtgagcagctcagcagagcaccagccttacatttggagaaggt 1200
 Db 1189 CTGAATTCGGAAGTGTGGAGCAGTCTACGAGAGCACCAGCCTTACATTGGAGAAGGT 1248
 QY 1201 gccgaatctccggggggcagagctgcgcgaggaagaaagaaagcgaaggtttccgcggg 1260
 Db 1249 GCGCAATCTCCAGGGGGCCAGAGCTGCCGCGGAGGAAAGAACCCGAAGGTTTCCGCGCGG 1308
 QY 1261 gagcggggtgcacacccagcctcgcgagagagagcccccgcggttaacaaagcagcg 1320
 Db 1309 GAGCCGGGTTCGACCCCGAGCCCTGCGCCGAGGAGAGGCCCGCGGTAAACAGAGCGG 1368
 QY 1321 gagcgagcgcctcgggcccccagggagagagagaggtcgttaagatagagacccag 1380
 Db 1369 GACGGAGCGCCTCGCGCCGCCACGGAGGAAGAGGAGGAGGTCTGTTAAGGATAGGACCCAG 1428
 QY 1381 cagagtaaacctcagcccccccgagcccaagcacaaggtcctcctgactggagctcctacc 1440
 Db 1429 CAGAGTAAACCTCAGCCCCCGCCGACCAAGCACAGGCTCCCTTGACTGGCAGTCTCACCCTT 1488

QY 1441 tcagagctgtctgtgaactcgcaggggggagtggtgcccctgcccacagtgaccttgytctccc 1500
Db 1489 TCCGAGCTGTCTGTGAACTCGCGAGGGGAGTGGCCCTGCCAACGTGACCTTGTCTCCC 1548
QY 1501 aaactgagccccagaccagagcctctcccttgaatcaagcccgctgtgaaatgaccag 1560
Db 1549 AACCTGAGCCCCGACACCAAGCAGGCGCTCTCCCTTGAATCAGCCCGCTGCTGAAATGACCAG 1608
QY 1561 gectgccccgagcagcagatgaggtgagggcgaggaagagattcccaactgataaa 1620
Db 1609 GCCTGCCCGGACGACGATGAGGATGAGGCGCGGAGGAAGAGATCCCAACTGATAAA 1668
QY 1621 gctactcaactaagaaagtgtctaccaccagcagcgaacatatctgaaaggtctctcgaa 1680
Db 1669 GCCTACTTCTATAGTAAGAAAGTGTCTACCACCAGCAGTACCAATATCTGAAGGATCTCGAA 1728
QY 1681 gttatcactctggtttcagagcagtgagcaagagagacccatgcccgaagcactg 1740
Db 1729 GTTATCATTCTGTGGTTTCAGACAGTGAAGCAAGAGGACCCCATGCGCGAAGCACTG 1788
QY 1741 aaagtctcatatcccgaaattttgaacctttgcacaaatttcatataattttctcaag 1800
Db 1789 AAAAGTCTCATATTCGGAATTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAG 1848
QY 1801 gaaattgagcaacgaacttgccctgtgtgggaagccgctcctcaaatgcccaaatcagagattac 1860
Db 1849 GAAATTGAGCAACGACTTGGCCCTGTGGGAAGGCGCGCTCAAAATGCCAAATTCAGAGATTAC 1908
QY 1861 caaagaatcgccatgctatgctgaagaacattcagggaatgaagacactgagcggtctcac 1920
Db 1909 CAAAGAATCGCCGATGTCACTGTGAAGAACATTCAGGGCATGAAGCACCCTGGCGGGCTCAC 1968
QY 1921 ctgtggaagcacagcagggccttggagccctggagaaatggaatcaagagctcccgccggy 1980
Db 1969 CTGTGAAGCACAGCGAGGCGCTTGGAGGCGCTGAGAGATGGAATGGAATCAAGAGCTCCCGCGG 2028
QY 1981 ctggagaactctgcagagactttgagctgcagaaagtgtgttacctaccgctcaacacc 2040
Db 2029 CTGGAGAATCTCTGCAGAGACTTTGAGCTGCAGAAAGTGTTGTACCTACCGCTCAACACC 2088
QY 2041 ttctctctcgccactgcaccggtctatgctactaagcagctcctgagcagctgtgc 2100
Db 2089 TTCTCTCGGGCCACTGCAACCGGCTCATGCACTAAGCAGGCTCTGGAGCGGCTGTGC 2148
QY 2101 aaacacacccgcgagccacgcgacttcaggagactgcgagccgctttggcagagatc 2160
Db 2149 AAACACACCCCGGAGCCGACGCGGACTTCAGGAGCTGCGGAGCGCTTTGGCAGAGATC 2208
QY 2161 acggagatggtgcacagctccacggtacgagatgatacaagatgagaaatttccagaaagctg 2220
Db 2209 ACGGAGATGTTGGCACAGCTTCCACGGTACCGATGATCAAGATGAGAAATTTCCAGAAAGCTG 2268
QY 2221 cagaaactcaagaatttgattgacattgacattgacatttctgttccggaaaggagttc 2280
Db 2269 CAGCAACTCAAGAAGATTGTATGGCATTGACAAATCTTGTGTTCCGGGAAGGGAGTTC 2328
QY 2281 atccgtctgggcagcctcaagagctctcgggaagggtccagcagcagctgttcttc 2340
Db 2329 ATCCGCTGGGCAGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTC 2388
QY 2341 ctgttcaacgaactctctgtatatacagagccgggggtgagcgccctcccaatcagtttaa 2400
Db 2389 CTGTTCAACGACGCTCTGTGTATACAGAGCGGGGGCTGACGGCTCCCAATCATGATTAAA 2448
QY 2401 gtccacggcagctcccgctctatgcatgacatgacatgagagagcgaagcagtgaggg 2460
Db 2449 GTCCACGGGAGCTCCCGCTCTATGCGATGACGATTGAGGAGAGCGAAGCAGTAGTGGGG 2508
QY 2461 gtgcccactgctgacctccggggccagcggcagtcacatcagtcgtggcccgagttct 2520
Db 2509 GTGCCCCACTGCTGACCTTCCGGGGCCAGCGGCGAGTCCATCATCTGTGGCGCCAGTTCT 2568

QY 2521 cgtccgagatgagaaagtgggttgaggacatccagatggccattgacctggcggagaag 2580
Db 2569 CGGTCCGAGATGAGAAGTGGGTGTGAGACATCCAGATGGCCATTGACCTGGCGGAGAAG 2628
QY 2581 agcagcagccccccctgagttccttgccagcagcagccccctgacaacaagtccccctgat 2640
Db 2629 AGCAGAGCCCCCCTGAGTTCTTGGCCAGCAGCCCCCTGACACAAAGTCCCCCTGAT 2688
QY 2641 gaagccacgcgggctgaccaggagtcagagagatgacctgagcgcctgcgcacatgcgtg 2700
Db 2689 GAAGCCACCGCGCTACCAAGGAGTCAGAGGATACCTGAGCGCTCGCGCACATCGCTG 2748
QY 2701 gacgcccagggccccgaccgcggaacacaaatggtgcaactgctgagccgcaacacc 2760
Db 2749 GAGCGCAGGCCCCCGCACCGCGCAACAATGTTGACCTGCTGGCACCGCAACCTG 2808
QY 2761 agcgtctccatggtgacttcagcatcgagtgaggaatcagttgctggaacactgctg 2820
Db 2809 AGGCTCTCCATGTTGACTTCAGCATCGCAGTGGAGAAATCAGTTGCTGGAAACCTGCTG 2868
QY 2821 agaaattcaaaaacagcaggggtggcagaagctgtggtgtgttcacaaaactttctgc 2880
Db 2869 AGGAAATTTCAAAAACAGCAACGGGTGGCAGAAGCTGTGGTGGTGTTCACAAAATTTCTGC 2928
QY 2881 ctgttcttcaaaaacacaccaggaacaatcatcccttccagcctgctgctgcgcg 2940
Db 2929 CTGTTCTTTACAAATACACACAGGACAATCATCCCCCTGCCAGCTGCTGCTCGGC 2988
QY 2941 tactcgtcaactccctctgagtcgagaaatccagaaagactacgtgttcaagctg 3000
Db 2989 TACTCGCTACCATCCCTCTGAGTCCGAGAAATCCAGAAAGACTACGTGTTCAAGCTG 3048
QY 3001 cacttcaagtccacgctctactacttcaggcgggaagcagagtacacgttctgaaagtgg 3060
Db 3049 CACTTCAAGTCCCAACGCTACTACTTCAAGCGCGGAAAGCGAGTACAGCTTCGAAAGGTGG 3108
QY 3061 atggaagtgcagcagtgccaccagctctgctcgcgacccacagctgttgagccacaaa 3120
Db 3109 ATGGAAGTGTATCCGCAAGTGCACACAGCTCTGCTCGGACCCACAGTGTGAGCCACAAA 3168
QY 3121 gagtctcttctgtat 3135
Db 3169 GAGTCTCTGTGTAT 3183

RESULT 2
BC004009 LOCUS BC004009 2632 bp mRNA ROD 12-JUL-2001
DEFINITION Mus musculus, clone IMAGE:3493093, mRNA, partial cds.
ACCESSION BC004009
VERSION BC004009.1 GI:13278387
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2632)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu

REFERENCE 2 (bases 1 to 3997)
 AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
 TITLE Direct Submission
 JOURNAL Submitted (08-OCT-1998) to the DBJ/EMBL/GenBank databases. Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES

source location/Qualifiers
 1. 3997

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hk05692"
 /clone_lib="pBluescriptII SK plus"
 /dev_stage="adult"
 /sex="male"
 /tissue_type="brain"
 118. .3282
 /gene="KIAA0793"
 118. .3282
 /gene="KIAA0793"
 /codon_start=1
 /product="KIAA0793 protein"
 /protein_id="BAA34513.1"
 /db_xref="GI:3882307"
 /translation="MGEIEGTVRVLTQAGRLGAQTPVGVSTLEPGQTLPRMQEKHL
 HLRKLDNTWEIFDIEPKDQVLLTQWRLNLECDYFCMEFQNTQSYWIWLEPM
 KPIIRLRPNVLRVLAFFPPDQGLQEQYRYLFALOLKRDLRLERLTCAADTTA
 ALLSHQLQSDIGYDETLDRHLKVNELYLPQOHCLEKILEFHOKHVGTQPAESDFQ
 VLETKRLMEYGIHFMAADREGTKIOLAVSHMGVLVFGQTKINTFNWVKRLSFK
 RKRLIKLHPVHGQYDTLEFLGSRDECKNFWKICVEYHTFFRLDQPKAKAVF
 FRSGSFRTGKQLVDFKDSMKRPIYERRSKHTHTSVRALTDLPKQISIFPE
 GLRTPASSANAFYSPLSPGLSPLEPKDSSLDQPVSYVKSPAAERRSGAVA
 GDTPTSAQGLPALQPLGTSKSPQSPSRKSPLSLSPAFQVPLGPAEQSSPL
 LSPVLSDAGMGECEEPHRKRVPADEAYFVKEILATERYLKDELYITVWFRSAV
 KEDAMPATLMLLSNIDPIYEFHGRFLREVEORLALWEGPSKATKGSHORIGDILL
 RNMQLKEFTSYFQRHDSVLELEKATRKCKLEAVYKEFELQVCIYLPNFTLLKPI
 QRLHYRLRLARCFYSPGHVDYADCHDLKAITEVTTTLOHILIRILENLOKULEQ
 LDVGIELNLRGHTFREGCLHLTKRGLQRMFFLFDMLLYTSKVAQTSFHRIR
 GLPLQGLMVEESNENSWPHCFITYAAQKTIVAASTRLEKEKWMLDLSAIOAKS
 GGDTPALPTGTVCTRPSPNEVSELESDDARGVRSLEGHGHRANTTMHVCWY
 RNTSVSRADHSAANVENOLSYLLRKFNKSHGOKLWVFTNFCFLFFYKTHODDYPLAS
 LPLIGSVSIFPREADGIHKQVFKLQFKSHVFFRAESKYTFERWMEVIQASSAGR
 APSVDQGPSPSSLEGVMVRGKEE"

BASE COUNT 998 a 1132 c 1068 g 799 t

Query Match 33.1%; Score 1037.4; DB 9; Length 3997;
 Best Local Similarity 60.2%; Pred. No. 2.9e-221;
 Matches 1864; Conservative 0; Mismatches 1186; Indels 45; Gaps 7;
 QY 21 gccacccagatcacgactggggccggccgaaattcggggatcagtcacctggaagc 80
 Db 150 GCAGACTGCGAGGATGCGCTTGGGTGCCAGACCCCTGTGGAGTTAGCACCCCTTGAGCC 209
 QY 81 tggacagaagcgcgcccaaccttcagaaaaactcgtgtccatcaaatccagatgct 140
 Db 210 TGGCAGACTCTCTTGGCCAGAAATCAAGAGAGACACCTGCACCTCAGAGTAAAGCTGCT 269
 QY 141 gaatgacacccaggaggaatttgaagttccacaaagagctccttggaagtgctgctgga 200
 Db 270 GGACACACCATGGAATAATTTGACATTGAGCCTAAATGCGATGCGCAGGTATTACTGAC 329
 QY 201 tgcagttgacacacccctcaactcgtggaagtgactattttggcctcgagtttccta 260
 Db 330 ACAAGTGTGGAAGCGTTTAAACCTGGTAGAATGTGACTACTTCGGGATGGAGTTTCAAAA 389
 QY 261 tcacaaaaagatacaggtgtggtggtgactcctcaaaacccactgtgaaacagattagaag 320
 Db 390 TACTCAGTCTTACTGATTTGGCTTGAACCTATGAACCCCATCATTTAGCAAAATACGAG 449
 QY 321 gccaaagcagctgtgttgaattgtgtggaattcttctccgctgacacacacaaact 380

Db 450 GCCAAGAATATGGTGTCTTCGCCTAGCTGTAAAATTTTTTCCACCTGATCTCTGGTCAGCT 509
 QY 381 ccaagaagaactcaacaagcttcgcgtcgcaggtggaagcagcagctgtgctcaagg 440
 Db 510 ACAAGAAGAATATACAAAGATACTTGTTCGCTTGCAACTTAAGAGAGACCTGCTGGAAGA 569
 QY 441 csaggtgacgtgtaataagcaccagcagcagctctctgtattcacacatgtgcaaatctga 500
 Db 570 GCGTTTGACCTGTGCTGACACACACAGCGGCCCTTCTTCAGTCCCTCTCTCGAGTCGA 629
 QY 501 gattgggagatttgatgaagccttgacagagagcacttagcaaaaaataatacatacc 560
 Db 630 AATAGAGATTACGATGAACCGCTGGACGACGAGACCTCAAAAGTGAACGAGTATTTGGC 589
 QY 561 tcagcaagacgcactagagacaaaatcgtggaatttcaccataaccacacttggaacaac 620
 Db 690 TGGCCAGCAGCAGCTGCTTTGAGAAGATACTAGAATTTCCATCAGAACGCTGGGCCAGAC 749
 QY 621 accagcagaatcagatttccagctcctagagattgcccgtcgcgtagagatgtatgaat 680
 Db 750 ACCTGCTGAGTCGGATTTCCAGTGTCTGAAAATTTGTCGAAAATTTGAAAATGTTACGGCAT 809
 QY 681 ccggttgaccgcggccaaagacaggaagcagcaagatcaatctcgcgttgcccaaac 740
 Db 810 CAGATTTACATGGCTTCTGACAGGAAGAACCAAGATTCAAACTGGCAGTTTCCCACAT 869
 QY 741 gggaaattcagtttccaggggttccactaagatcaatgccttcaactggcggcgaagtgcg 800
 Db 870 GGGTACTCGTGTCCAGGGCACCACCAAAATCAACACTTTCACTGGTCCCAAGTCCG 929
 QY 801 gaagctgagcttcaagaggaagcgtcttctcatcaagctccgcccagatgcgaatagtc 860
 Db 930 TAAACTAAGCTTCAAGAGCAAAAGATTTCTTATCAAACTTCATCCAGAGGTTTCATGCACC 989
 QY 861 gtaccagatacttggaattcctaagtgccagtcgggattctcgaagtctcttgaa 920
 Db 990 TTACCAGGACACATTAGAATTTTGTGGGTAGTAGAGATGAATTAAGAACTCTTGGAA 1049
 QY 921 aatctgtgtggaacacatcgtcttcttagacttttgaagcggcccaaaccaagcccaa 980
 Db 1050 GATTGTGTGGAGTATCACACCTTTTGTAGACTTTTGGACCAACTTAAGCCAAACAAA 1109
 QY 981 gccgctcctcttagcgggggttcattctcgtgttcagtggtcggactcagaagcaggt 1040
 Db 1110 AGCGGTCTTCTTCAGCGGGGCTCCCTCTCAGATACAGTGGGAAGAACTCAGAAACAACT 1169
 QY 1041 tctcgaactgttaaaagaaggagacataaagaaggtgcagtttgaaaggaagcacagcaa 1100
 Db 1170 AGTAGATTATTTCAAAGACAGTGGAAATGAAGAGAATTTCCATATGAAGAAGGCACAGCAA 1229
 QY 1101 gattcatttatccggagccttgcttcacagcctcacagaactgaattcggaaagtgcctgga 1160
 Db 1230 GACCCACAGTGC-----GTTCCAGGCTCTGACTGCAGACTTACCAA 1271
 QY 1161 gcagttcagcagagcaccagccttacatttgagagagtgccgaatctccagggggcca 1220
 Db 1272 ACAGAGACTCTCATTCGCCGAGGGATTGAGGACTCTCTGCTTCCCATCTTTCAGCGAATGC 1331
 QY 1221 gggctgcggcgagaaagaaacccgaaggtttccgcgggggagcggggtcgcaccccgag 1280
 Db 1332 CTTTACTCGCTCTCTCCCTCCACTCTGTGCCCTCTGTGCCCTGCCAGAGTTTAAAGACAG 1391
 QY 1281 cctcgcgcgagagagaagcccgcggttaacaagcagcgcgagcgcgcgcctcgcgcgc 1340
 Db 1392 CAGCAGCTCCCTCAGAGATCCCGAGTTTCTTACGTCAAGAGTCCAGCTGCAGAGAGGGG 1451
 QY 1341 cacggagggaaggagagag---tcgttaagatagagaccacagcagagataaacctcagcc 1397
 Db 1452 CAGTGAGCAGTGGCTGGAGGCCCGCCGACACACCATCGGCCGACCCCTCGGGCCCCCGC 1511
 QY 1398 ccgcagcccaacacagaggtccctgagtcagtcctacccttccgagctgtctgtgaa 1457

Db 1512 ACTCCAGCGTGGTCCAGGCGCTTTTCCACGAAGAGTCTCTCAGCCTTCTCCCTCCAGCGGAA 1571
Qy 1458 ctgcaggggggagtgagccctgccaacgtgacctgtctctccaaacctgagccccgacac 1517
Db 1572 GAGCCCCCTGAGTCTGAGCCCTGCTATTTACAGTGCCTT-----TGGGCCACAGCTGA 1622
Qy 1518 caagcaggcctctcccttgatcagccgcgtgctgaaagacagcagcctgccccggagcga 1577
Db 1623 ACAGGGCTATCCCACTCTCTGAGCCCTCTCTCAGTATGCTGGCGAGGCCGGGATGGA 1682
Qy 1578 cgaatgagatgagggccggaggagagatctcccaactgataaagcgtactctatagctaa 1637
Db 1683 CTGGAG---GAGCCACAGACAAACAGCGCTGCTGCACAGAGGCCCTACTTCATAGTCAA 1739
Qy 1638 ggaagtctaccacagcagcaacatactgaagatctcgaagatctcgaagtataactcgtggtt 1697
Db 1740 AGAGATTCTCGCTACAGAACCAATACCTCAAGGATTTAGAAATTTATACCTGTGTGTT 1799
Qy 1698 tcagagcacagtgagcaaaagagcgcctgcccgggaagcactgaaaagtctcatattccc 1757
Db 1800 CCGCAGCGCAGTGGTGAAGGAGCGCCATGCTGCGACTCTGTATGACGCTGCTCTCTC 1859
Qy 1758 gaattttgaacctttgcacaaatttoatactaatattttctcaagaaattgagcaagact 1817
Db 1860 CAACATCGATCCCATCTATGAGTTCCACAGAGGCTTCTCGCGAGGTGGAGCAGAGGCT 1919
Qy 1818 tgccctgtggaaggcgtcaaatgcccaaatcagag---attacaaagaatcggcga 1874
Db 1920 GGCACCTGGGAAGGGCCCTTCAAGGCCACACAAAGGAGCGAGTATCAACGAATCGGGGA 1979
Qy 1875 tgcctatgaaagaacattcagggcataagcaccttgccggtcacctgtggaagcacag 1934
Db 1980 CATCTGCTCAGGACATGCGCCAGTTAAAGGAGTTTACCAGCTACTTCCAAAGACATGA 2039
Qy 1935 cgaggccttgaggccctggagaatggaatcaagagctcccgcggtggagaaactctg 1994
Db 2040 CGAGGCTCTAACGAACCTGGAAGGCTTACCAACGCTGTAAAGTTGGAGCAGTGA 2099
Qy 1995 cagagacttgagtcagagaagtgtgtacctacgctcaacacctctcctgagcc 2054
Db 2100 CAGGAGTTTGTAGCTGAGAAGGTCTGCTACTTGCCTCTCAACACGTTCTGCTGAAGCC 2159
Qy 2055 actgcacggctcatgcactacaagcaggtccctggagcggctgtgcaaacaccaccgcc 2114
Db 2160 CATCCAGCGCTGCTGCACATACCCCTGCTGCTGCGCGCTATGCGGACATATACAGCCC 2219
Qy 2115 gagccacgcgaacttaaggactgcgagccgcttcttgacagatacagagatggtgac 2174
Db 2220 CGGGCACCATGACTACGCTGACTGCCATGACGCCCTTGAAGGCCATCACAGAGGTGACCAC 2279
Qy 2175 acagctccacggtacgatgatcaagatggagaatttccagaagctgcagaaactcaaga 2234
Db 2280 CACACTACACACATCTCATCGGCTGGAGAACCTGCAAGACTACGAGGTGACGCG 2339
Qy 2235 agatttgattggcattgacaatcttggttccgggaaggaggttcatcctcgttggcag 2294
Db 2340 GGACCTGGTGGCATAGAGAACCTCATTTGCTCTGCGAGGAGTTTCATCGTGAGGCGTG 2399
Qy 2295 cctcagaacactctcggggaagggttccagcagcagcatgttctcctgttcaacagact 2354
Db 2400 CTTTCACAACTCACCAGAAGGCCCTGCGACGAGAGGATGTTTTTTCTGTTCTCAGATAT 2459
Qy 2355 cctgtatcacagcagcggggtgctgacgacctccaatcagtttaagtccacggcagct 2414
Db 2460 GTTGCTGTACACAAAGAGTTGAGGAGCACCGCCATCTCCGATCCGGGCGCTCTCT 2519
Qy 2415 cccgctctatggctgacgattgagagagcgaagcagtgagggtgggggtccccactgcct 2474
Db 2520 TCCCTTCAAGGCATGCTGTGTGAAGAAAGTATACGAGTGTGCTTCTCCACACTGTTT 2579
Qy 2475 gaccttcggggccagcggcagtcctatcctgctggtggccgagctctcgttccagatgga 2534
Db 2580 CACCATCTACGCGCTCAGAAAAAATCGTGTGGTGGCAGCCAGCAGCTCGGCTGGAGAAGA 2639

Qy 2535 gaagtgggttgagagaccatccagatggccattgacctggcgagagaagagcagcagccccgc 2594
Db 2640 GAATGGATGCTGGACCTGAACCTCCGCTGAACCTCCGAGTCCAGCAAGAGTGGGTGACAGGC 2699
Qy 2595 cctgagttcctggcagca-----gcccccctgacaacaagtccccctgatgaagccac 2648
Db 2700 CCTGTGACTGCCAGCGCGACTGTGTGCACTGTCTCCCGCCAGATCCCGCAACAGGATATC 2759
Qy 2649 cgcggtgaccagagtgacaggtgacgatgacctgagcgccctgcgcacatcgcgtgagcgcca 2708
Db 2760 TCTG---GAGCAGAGAGTCAGAAGATGATGCTCGGGGTGTCCGAGCTCCCTGGAGGGCA 2816
Qy 2709 gccccgcacccgcgcgaacaacaatggtgcacgtgtgctgcacccgcgaacaaccagcgtctc 2768
Db 2817 TGGCCAGCAGCCGGGCCAACACCACTCAATGACAGTGTGTGTACCGAACACACGCGTGC 2876
Qy 2769 catggtgacctcagcatcagtgaggaatcagtgctggaacacctgctgaggaatc 2828
Db 2877 CAGGCAGACCCACAGTGCACCTGTCGAGAACACAGCTTTCAGGATATCTGCTAAGAAAGTT 2936
Qy 2829 caaaacagcaacgggtggcagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2888
Db 2937 CAAAAACAGTCATGCTGGTGGCAGAAGCTCTGGTGTCTTTTACCAACTCTCTGTGTCTT 2996
Qy 2889 ctacaaatcacacacagacaatcatcccttgcagcctcctcctcctcctcctcctcctcct 2948
Db 2997 CTAAAACTCATCAGGTGACTACCCACTGGCCAGCTCCCGTGTGGGTATACACGCT 3056
Qy 2949 caccatccccctctgagtcgcgagaacatcccaagaagactcgtgttcaagctgcactcaa 3008
Db 3057 GAGCATCCCCAGGAGCGCGATGSCATACACAAAGACTATGTTTCAAGCTCCAGTTCAA 3116
Qy 3009 gtcccagctctactacttcaggcggaagacgcagagtgacacgttcgaaaggtcggaagt 3068
Db 3117 ATCCACGCTACTTCTTCCGGCTGAGAGCAAGTACACATTTGAAAGGTGGAGGT 3176
Qy 3069 gaccgagtcgaccagctcctcctcgacccc 3103
Db 3177 GATCCAGGGGCCAGCAGCTCAGCGGAGGGGCC 3211

RESULT 4

BC009153 3719 bp mRNA ROD 12-JUL-2001
LOCUS Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304
DEFINITION IMAGE:2655209, mRNA, complete cds.
ACCESSION BC009153
VERSION BC009153.1 GI:14318718
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3719)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 5 Row: m Column: 20.

FEATURES

Location/Qualifiers

source

1..3719
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:6304 IMAGE:2655209"
/tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
99..3296

CDS

/codon_start=1
/product="Similar to KIAA0793 gene product"
/protein_id="AAH09153.1"
/db_xref="GI:14318719"
/translation="MGEIEGYRALPTSGTRIGGQTAIGVSTLEPEQSLSPRMQEKHM
RIRVRLDSTVEFLRDEPKDCQVLLTQWVKHLNIECDYFGLFKNVSYWIMLEPM
KPIRQVKNPNAVLRAVKFPDPDGOLOEYTRYLALQIKRDLERLCTANTTA
ALLISHLQSEIGDYDEFLDREHLKANEYLNQESLKEILDFHQHRTGQTPAESDFQ
VLEIARKLEMGIIRFHMSADREGTKINLAVSHMGLVFGQTKINFWNSKVRKLUKFK
RKRFLIKLHPVHGYPQDTLEFLLSGDECKNFKICVIEYHTFFRLSDQPKAKARF
FLRSTSYRSRNTQQLVDYVKGDMKRIPLYRRHSKRTSLHALTVLDPKQSVFTD
GIRGSLSSANVSFPYPPSPSLSPGLNPKLSDSSSLVDPQAPVIKGTAABRSRSGPS
SSDGPSTOAHLPGLPVRPGFSDPSPSPSLKSHLSICPELOALSPAEQOGAS
PVLSVLSGACTARMNDNEQKHMPDEAYFAKEILATERYILKDEIVTYWFRS
VLKIEEAPMALNLFIDPVEFHFGFLEVEQRLMEGSPSAHLKHQDHQIGD
ILLNRMLQKFTSYFQRHDEYFLELEKATCKKLEAYKEFELOKVCYLPPLNTFL
KPVQLRYRLLSRLCAHSYSPHRDYADHEALKAITVETTELQOISLRLENLAKLT
ELORDLVGNELIAPREFIREGDLKLTKKGLQORMFLFSDMLLYTSKSVTGAHF
RIRGPLRLGLVPERSENEWSVLHCFITYAAQKTVAASTLEKEKMQDLNAAIOA
AKTGDSPVLLGGVRYTRPSRSEVSESEDEGRNGRSLGNSQHRANTTMVWC
YNTSVSRADHSAVENSGLTLLRFRKNSNGWKLVVTFNCLFFFKTHODDYLPLA
SLPLGYSVLPREADSTHKDYFKLQKSHVYFFRAESKYTFERWMDVIRASSSPG
RPPSTQDCSHHSGLEAIREKEACPSPCLDKNL"

BASE COUNT 997 a 979 c 929 g 814 t

ORIGIN

Query Match 32.3%; Score 1013.2; DB 10; Length 3719;
Best Local Similarity 59.3%; Pred. No. 7.3e-216;
Matches 1846; Conservative 0; Mismatches 1238; Indels 30; Gaps 6;
QY 22 ccgacccagatcacgactggggcccggaataatcggggatacagtaacctggaaagt 81
DB 132 CCAACTTCAGGACACGCTTGGTGGCCAGACCGCCATTGGAGTGAGTACCTCGAGCCA 191
QY 82 ggacagaagccgccccacacattcaggaaactcgtgtccataaaatccagatgctg 141
DB 192 GACGACGCTGTACACAGGATGACGAGAGACACATCGCATCAGATGAAGCTGCTG 251
QY 142 gatgacacccaggagcattgaagtccacaaagagctcctgggaagtgctgctggat 201
DB 252 GACAGCACTGTAGAGTATTGTGATCGAGCGCAAGTGTGACGGGAGCTTACTGACT 311
QY 202 gcagtttgcaaccactcaactcgtggaggtgactattttggcctcagatttctctgat 261
DB 312 CAAGTGTGAAGCAATTTGAATTCATTTGAATGTGACTTACTTGTGGTGTCAAGAA 371
QY 262 cacaaaaagatcacggtgctggtatcctctaaacccatgtgaacagattagaagg 321
DB 372 GTCCACTCTACTGGAATTTGGCTTGAACTATGAAACCCATCATATAGGCAAGTACGAA 431
QY 322 ccaagcagcgtgtgttaagtttggtgaaattcttcgcgctgacacacacacac 381
DB 432 CCAAAAAATGCGGTGCTTCGCTGGCAGTAAATTTTCCCGCCTGACCTGCTGACGCTG 491
QY 382 caagaagaactcacaaggtaacctgttctcgctgcaggtgaagcaggaacttgctcaaggc 441

[illegible]

```
QY 2845 tggcagaagctgtgggtgttcacaaactctgctgtcttcttcacaaatcacaccag 2904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 365 TGGCAGAAGCTCGGGTCGCTTTACCAACTCTGTTGTTCTTCTACAAAACATCATCAG 424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2905 gacaatcatcccttgccagcctgctctgctcggtactactcgctaccatccctctgag 2964
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 GATGACTACCACTGCGCCAGCCCTCCCGCTGCTGGGTACAGCGTGAGCATCCCGAGGAG 484
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2965 tcgcagaacatcagaaagactcgtgtcaagctgcacttcaagtcaccagcttactac 3024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 485 GCGGATGGCATACAAAGACTATGTTTCAAGCTCCAGTTCAAATCCACGCTCTACTTC 544
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3025 ttcaggcggaagcagtagtacacgttcgaaagtgatggaagtgcgcagctgcacc 3084
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 TTCCGGGCTGAGACGAAGTACACATTGAAAGGTGATGGAGTGATCCAGGGGCCAGC 604
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3085 agctctgctcgcgacccc 3103
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 605 AGCTCAGCGGAGGGCCC 623
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AL161896 AL161896 96183 bp DNA PRI 20-JAN-2001
LOCUS Human DNA sequence from clone RP11-261P24 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL161896
VERSION AL161896.16 GI:12330752
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 96183)
Phillimore,B.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:12329254.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-261P24 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTORT: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-261P24 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-261P24 is at 96183 in this
sequence. The true left end of clone RP11-31K22 is at 59428 in this
sequence. The true right end of clone RP11-573N10 is at 100 in this
sequence.
FEATURES
source Location/Qualifiers
1..96183
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-261P24"
/clone_lib="RPCI-11.1"
1301..1335
/note="SN5 repeat: matches 34. .68 of consensus"
1432..1506
/note="MER20 repeat: matches 1. .75 of consensus"
3518..3854
/note="L1MB8 repeat: matches 5837. .6173 of consensus"
5528..5599
/note="12 copies 6 mer tattta 69% conserved"
5533..5604
/note="36 copies 2 mer at 69% conserved"
5766..6068
/note="AluSp repeat: matches 7. .303 of consensus"
6949..7264
/note="AluJo repeat: matches 1. .307 of consensus"
7380..7501
/note="MIR repeat: matches 65. .191 of consensus"
8192..8407
/note="L2 repeat: matches 2064. .2289 of consensus"
8475..8506
/note="16 copies 2 mer tt 100% conserved"
8929..9103
/note="MER33 repeat: matches 4. .171 of consensus"
9104..9421
/note="AluJb repeat: matches 1. .306 of consensus"
9422..9450
/note="MER33 repeat: matches 171. .216 of consensus"
9491..9595
/note="MER33 repeat: matches 142. .225 of consensus"
9596..9909
/note="AluY repeat: matches 1. .311 of consensus"
9910..9997
/note="MER33 repeat: matches 225. .319 of consensus"
11665..11763
/note="MIR repeat: matches 47. .149 of consensus"
11991..12022
/note="8 copies 4 mer tgtg 90% conserved"
12274..12568
/note="AluX repeat: matches 20. .311 of consensus"
12577..12879
/note="AluY repeat: matches 1. .304 of consensus"
13406..13549
/note="MLT1C repeat: matches 1. .146 of consensus"
13554..13585
/note="16 copies 2 mer ca 96% conserved"
13682..13715
/note="17 copies 2 mer ac 97% conserved"
13684..13715
/note="8 copies 4 mer acac 100% conserved"
13752..13779
/note="14 copies 2 mer ca 96% conserved"
13784..14048
/note="MLT1C repeat: matches 189. .465 of consensus"
14716..15018
/note="AluSc repeat: matches 1. .303 of consensus"
15567..15842
/note="AluJo repeat: matches 3. .305 of consensus"
16752..16916
/note="MIR repeat: matches 63. .227 of consensus"
17700..17838
/note="AluJo/FRAM repeat: matches 163. .301 of consensus"
18111..18314
/note="MLT1A2 repeat: matches 157. .359 of consensus"
18317..18428
/note="28 copies 4 mer ctcc 71% conserved"
18319..18422
/note="52 copies 2 mer ct 63% conserved"
18322..18423
/note="17 copies 6 mer tctctc 63% conserved"
```

repeat_region 18510..18797
/note="AluJo repeat: matches 1. .302 of consensus"
repeat_region 18800..18956
/note="MLTIA2 repeat: matches 1. .171 of consensus"
repeat_region 19392..19704
/note="AluJo repeat: matches 1. .312 of consensus"
repeat_region 19780..20069
/note="AluJ repeat: matches 1. .290 of consensus"
repeat_region 22285..22456
/note="MIR repeat: matches 23. .220 of consensus"
repeat_region 22457..22760
/note="AluSg repeat: matches 1. .304 of consensus"
repeat_region 22761..22773
/note="MIR repeat: matches 220. .230 of consensus"
repeat_region 26734..26961
/note="L1MB2 repeat: matches 5944. .6170 of consensus"
repeat_region 28074..28174
/note="MER44C repeat: matches 7. .103 of consensus"
repeat_region 28175..28840
/note="MER72 repeat: matches 5. .668 of consensus"
repeat_region 28841..29431
/note="MER44C repeat: matches 103. .721 of consensus"
repeat_region 29821..30105
/note="AluSx repeat: matches 1. .285 of consensus"
repeat_region 30219..30775
/note="L2 repeat: matches 2150. .2749 of consensus"
repeat_region 30924..31207
/note="AluSx repeat: matches 3. .294 of consensus"
repeat_region 32053..32298
/note="MIR repeat: matches 1. .257 of consensus"
repeat_region 32498..32740
/note="L1MB3 repeat: matches 5943. .6182 of consensus"
repeat_region 33867..33920
/note="27 copies 2 mer ca 90% conserved"
repeat_region 33871..33918
/note="8 copies 6 mer cacaca 93% conserved"
repeat_region 33878..33921
/note="11 copies 4 mer acac 97% conserved"
repeat_region 34367..34420
/note="9 copies 6 mer ctcttc 75% conserved"
repeat_region 34458..34576
/note="MIR repeat: matches 94. .219 of consensus"
repeat_region 34914..35006
/note="MIR repeat: matches 150. .242 of consensus"
repeat_region 35044..35414
/note="MLT1B repeat: matches 1. .386 of consensus"
repeat_region 35440..35480
/note="L2 repeat: matches 2423. .2464 of consensus"
repeat_region 35653..36028
/note="THE1B repeat: matches 1. .359 of consensus"
repeat_region 36683..37012
/note="AluSx repeat: matches 1. .308 of consensus"
repeat_region 37526..37818
/note="AluJo repeat: matches 1. .290 of consensus"
repeat_region 37957..38028
/note="36 copies 2 mer tg 86% conserved"
repeat_region 37968..38031
/note="16 copies 4 mer gtgt 90% conserved"
repeat_region 37980..38027
/note="8 copies 6 mer gtgtgt 100% conserved"
repeat_region 38553..39205
/note="L1MB3 repeat: matches 5545. .6183 of consensus"
repeat_region 39561..39944
/note="MSTA repeat: matches 1. .426 of consensus"
repeat_region 40827..41086
/note="AluSx repeat: matches 1. .259 of consensus"
repeat_region 41793..42089
/note="AluSg repeat: matches 1. .297 of consensus"
repeat_region 42330..42393
/note="L2 repeat: matches 2627. .2701 of consensus"
repeat_region 42692..42857
/note="L1MC2 repeat: matches 5654. .5816 of consensus"
repeat_region 42858..43151

/note="AluSg repeat: matches 1. .293 of consensus"
43152..43183
/note="L1MC2 repeat: matches 5816. .5846 of consensus"
43184..43479
Query Match 8.1%; Score 254; DB 9; Length 96183;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1162 cagtcacagcagcaccagccttacatttgagaggtgcccgaatctccagggggccag 1221
|||||
Db 95665 CAGTCTCAGCAGCAGCAGCCCTTACATTTGGAGAAGGTGCCGAATCTCCAGGGGGCCAG 95724
QY 1222 acgtccggcgaggaaagaaacgaaggtttccgcggggagccgggtcgaccaccgagc 1281
|||||
Db 95725 AGCTCCGCGCAGGAGAAAGAACCGAAGGTTTCCGCGGGAGCGGGGTGCGACCCGAGC 95784
QY 1282 cctgcgcgcagagaaagccccgcggtaacagcagcggagcggagcgcctcgcgcc 1341
|||||
Db 95785 CCTGCGCCGAGGAGAGCCCCCGGGTAAACAAGCAGCGAGCGGCCCTCGGGGCC 95844
QY 1342 acggaggaagcagcagcaggtcgttaagcagcagcagcagcagcagcagcagcagc 1401
|||||
Db 95845 ACGGAGGAGAGGAGGAGGTGCTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCG 95904
QY 1402 cagccaagcacagg 1415
|||||
Db 95905 CAGCCAAGCACAGG 95918
RESULT 8
AC022669/c
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-96B23 map 15, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
ACCESSION AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 152053)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-96B23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 152053)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,K.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:7229803.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

Center project name: L4938
Center clone name: 96_B_23

----- Summary Statistics

Sequencing vector: M13; M77815; 93% of reads
Sequencing vector: Plasmid; n/a; %0.f% of reads
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136451 bases at least Q40

Consensus quality: 143612 bases at least Q30

Consensus quality: 146365 bases at least Q20

Insert size: 158000; agarose-fp

Insert size: 148853; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 784: contig of 784 bp in length
785 884: gap of 100 bp
885 1956: contig of 1072 bp in length
1957 2056: gap of 100 bp
2057 3113: contig of 1057 bp in length
3114 3213: gap of 100 bp
3214 4817: contig of 1604 bp in length
4818 4917: gap of 100 bp
4918 6582: contig of 1665 bp in length
6583 6682: gap of 100 bp
6683 8059: contig of 1377 bp in length
8060 8159: gap of 100 bp
8160 22369: contig of 14210 bp in length
22370 22469: gap of 100 bp
22470 24974: contig of 2505 bp in length
24975 25074: gap of 100 bp
25075 27206: contig of 2132 bp in length
27207 27306: gap of 100 bp
27307 29420: contig of 2114 bp in length
29421 29520: gap of 100 bp
29521 32000: contig of 2480 bp in length
32001 32100: gap of 100 bp
32101 36330: contig of 4230 bp in length
36331 36430: gap of 100 bp
36431 39067: contig of 2637 bp in length
39068 39167: gap of 100 bp
39168 42564: contig of 3397 bp in length
42565 42664: gap of 100 bp
42665 46664: contig of 4000 bp in length
46665 46764: gap of 100 bp
46765 51047: contig of 4283 bp in length
51048 51147: gap of 100 bp
51148 55269: contig of 4122 bp in length
55270 55369: gap of 100 bp
55370 58831: contig of 3462 bp in length
58832 58931: gap of 100 bp
58932 64077: contig of 5146 bp in length
64078 64177: gap of 100 bp
64178 68808: contig of 4631 bp in length
68809 68908: gap of 100 bp
68909 74170: contig of 5262 bp in length
74171 74270: gap of 100 bp
74271 79296: contig of 5026 bp in length
79297 79396: gap of 100 bp
79397 84724: contig of 5328 bp in length
84725 84824: gap of 100 bp

* 84825 89713: contig of 4889 bp in length
89714 89813: gap of 100 bp
89814 94721: contig of 4908 bp in length
94722 94821: gap of 100 bp
94822 100196: contig of 5375 bp in length
100197 100296: gap of 100 bp
100297 107015: contig of 6719 bp in length
107016 107115: gap of 100 bp
107116 114507: contig of 7392 bp in length
114508 114607: gap of 100 bp
114608 122671: contig of 8084 bp in length
122672 122771: gap of 100 bp
122772 128052: contig of 5281 bp in length
128053 128152: gap of 100 bp
128153 138583: contig of 10431 bp in length
138584 138683: gap of 100 bp
138684 150248: contig of 11565 bp in length
150249 150348: gap of 100 bp
150349 152053: contig of 1705 bp in length.
FEATURES
Location/Qualifiers
1. .152053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone_lib="RPC1-11 Human Male BAC"
1. .784
/note="assembly_fragment"
clone_end:SP6
vector_side:left
885. .1956
/note="assembly_fragment"
2057. 3113
/note="assembly_fragment"
3214. .4817
/note="assembly_fragment"
4918. .6582
/note="assembly_fragment"
6683. 8059
/note="assembly_fragment"
8160. .22369
/note="assembly_fragment"
22470. .24974
/note="assembly_fragment"
25075. .27206
/note="assembly_fragment"
27307. .29420
/note="assembly_fragment"
29521. .32000
/note="assembly_fragment"
32101. .36330
/note="assembly_fragment"
36431. .39067
/note="assembly_fragment"
39168. .42564
/note="assembly_fragment"
42665. .46664
/note="assembly_fragment"
46765. .51047
/note="assembly_fragment"
51148. .55269
/note="assembly_fragment"
55370. .58831
/note="assembly_fragment"
58932. .64077
/note="assembly_fragment"
64178. .68808
/note="assembly_fragment"
68909. .74170
/note="assembly_fragment"
74271. .79296
/note="assembly_fragment"

Insert size: 174530; 1.8% error; agarose-fp
Quality coverage: 6.26x in Q20 bases; sum-of-contigs Quality coverage: 6.72x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 114302: contig of 114302 bp in length
* 114303 114402: gap of 100 bp
* 114403 135098: contig of 20696 bp in length
* 135099 135198: gap of 100 bp
* 135199 173346: contig of 38148 bp in length
* 173347 173446: gap of 100 bp
* 173447 181800: contig of 8354 bp in length.

FEATURES
Location/Qualifiers
source
1. .181800
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q31.3-32.3"
/clone="RP11-111L24"
/clone_1lb="RPC1-11.1"
1. .114302
/note="assembly_fragment:00010
fragment_chain:1
clone_end:SP6
vector_side:left"
114403. .135098
/note="assembly_fragment:01859
fragment_chain:1"
135199. .173346
/note="assembly_fragment:02403
fragment_chain:1"
173447. .181800
/note="assembly_fragment:02166
fragment_chain:1
clone_end:T7
vector_side:right"
BASE COUNT 49780 a 43544 c 41179 g 46996 t 301 others
ORIGIN

Query Match 8.1%; Score 254; DB 2; Length 181800;
Best Local Similarity 100.0%; Pred. No. 2.1e-46;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1162 cagtctcagcagaccagccttacatttgagaaagtgccgagggggccag 1221
Db 36237 CAGTCTCAGCAGACACAGCCTTACATTTGGAGAAGTCCGAAATCTCCAGGGGCCAG 36296

Qy 1222 agctgcgcgcgaggaagaaaccgaaggtttccgcggggagccggggtcgaccccgagc 1281
Db 36297 AGCTGCCGCCGAGGAAGAACCGAAGGTTCCTCCCGGGGAGCGGGGTGCGACCCGAGC 36356

Qy 1282 cctgcgcgcgaggaagccccgcgcgggttaacaagcagcggagcggagccgctcgagccccc 1341
Db 36357 CCTGCCCGGAGGAGAGCCCGCGGGTAAACAAGCAGCGGAGGAGCGGCTCGCGGCC 36416

Qy 1342 acgaggaagaggaggaggttcgttaagataggaccagcagagataaacctcagccccc 1401
Db 36417 ACGGAGGAAGAGGAGGAGGTTCGTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCC 36476

Qy 1402 cagccaagcacagg 1415
Db 36477 CAGCCAAGCACAGG 36490

RESULT 10

Qy	2970	gaacatccagaaagactacgtgttccaagctgcactccaagtccaagctctactacttcaag	3029
Db	294	GAACATCCAGAAAGACTACGTGTTCAGCTGCACCTCAAGTGCCTCAAGTCCACGCTACTACTTCAAG	235
Qy	3030	ggcgaagaagcaggtacacgttcgaaaggtggatggaagtgatccgacgtgccaccacgctc	3089
Db	234	GGCGAAAGCAGGTACACGTTCGAAAGGTGATGGAAGTGATCGCGAGTGCACACGCTC	175
Qy	3030	tgctcgcgacccacacgtgttgagccacaaagagctctcttctgtat	3135
Db	174	TGCTCGGACCCACAGTGTGAGCCACAAAGAGTCTCTGTGTAT	129
RESULT	11		
AF156225			
LOCUS	AF156225	2595 bp	mRNA PRI 04-JUL-1999
DEFINITION	Homo sapiens protein 4.1 (EPB4.1)	mrna	complete cds.
ACCESSION	AF156225		
VERSION	AF156225.1	GI:5353737	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2595)		
JOURNAL	Huang, S.C., Wang, C., Lichtenauer, U., Vortmeyer, A. and Zhuang, Z.		
REFERENCE	Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 2595)		
JOURNAL	Huang, S.C., Wang, C., Lichtenauer, U., Vortmeyer, A. and Zhuang, Z.		
REFERENCE	Direct Submission		
TITLE	Submitted (03-JUN-1999) NCI, NIH, 9000 Rockville Pike, Bethesda, MD		
JOURNAL	20892, USA		
FEATURES	Location/Qualifiers		
source	1..2595		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="1"		
	/map="lp36"		
	/cell_line="LAN5"		
gene	/cell_type="neuroblastoma"		
	1..2595		
	/gene="EPB4.1"		
CDS	1..2595		
	/gene="EPB4.1"		
	/codon_start=1		
	/product="protein 4.1"		
	/protein_id="AAD4222.1"		
	/db_xref="GI:5353738"		
	/translation="MTKSLVTEAENSQHOKEEGEFAINSQQPEPQPEESCOTAAE		
	GDNCKBHLKASNGDPTDIEDLTKNKERTSENKGLRSLFSSFLKRPKQSVSEEGKEY		
	EDSKKEGGQKDEIFGTSLDEIEILKAPIAAPEPELKDPSLDLHLSLSAEVQPAQAE		
	ELRDEPDFETKEGGELEKCSIKVEKSPQSKAEATKLSQKP IRKRRNMHCXVLLD		
	DTVYCVKVKHGAQDILLKRVCCNLLLEDYFGLAIWNTTSTKTWLDNAKELTKVQR		
	GVPWNFTFNVKYVPPDAQLTETRYILCLQRODIVAGRLPCSPATALLGSGYTIQ		
	SELGDDYDPELHGVDDYVSDFKLANPQTKLEEKVWELHKYSRSTPAQADLEFLFNKAK		
	LSMVGDLHKGVDLDEVDIILGVCSSGLLVIKDKLRNRPFPMPVKLKI STKRSFFIK		
	IRPGEQVESTTGFKLPSYRAAKKLWCKVRSHTTFRUTSTDTIPKSFALGSKFR		
	YSGRTQAOTQASALDRPAPHPERTASRRSLDGAAVDSADRSPTSPATITQ		
	GOVAGEQDQKLAKEETDILRMBKRRRLDGENIYIRHSSLMLEDLKSOREIEKH		
	VPTNSGDQTKAAETDILRMBKRRRLDGENIYIRHSSLMLEDLKSOREIEKH		
	ASISLKNFMESVPEPRSEWDKRLSTHSPTFLNIHQIPTGEGPPLVKTQTVTIS		
	DANAVKSELPKIDVPIVHTETKTIYEAQDNTDNGSDLDGVLITQAQITSETSPST		
	TTNQITTKVKGISRTIRKIEKRLITGDADIDHDQVLVQAIKEAKEQHPDMSVTKVMVV		

QY 117 cgtgtccatcaaaatccagatgctggatgacacaccacccaggagcgatttgaagttccacaagaag 176
Db 627 CATGCACTGCAAGGTTCTTTGTTGGATGACACAGTTTATGAATGTTGTTGGAGAAACA 686
QY 177 agctcctgggaagtgctgctggatgcagtttgcacaccctcaacctcggtggaaggtga 236
Db 687 TGCATAGGCACAGATTGCTTAACAGAGTAGTGTGAGCATCTCAATCTTTTGGAGAAACA 746
QY 237 ctattttggcctcgagtttccctgatcacaaaaagatcacggtgtggtggtgatctctcaaa 296
Db 747 CTATTTTGGTCTAGCCATTGGGATTAACACAACCTCTAAGACATGCTGGATTCCGCCCAA 806
QY 297 acccatttgaaacagattagaagcccaagcacggtgtgttgaagtttgttggtgaatt 356
Db 807 AGAAATAAAAAGCAGGTTGCGTGG----TCTCCCTTGGAAATTTTACATTTTAATGTAAAGTT 863
QY 357 ctttcgcctgaccacacacaaactccaaagaactcaagaagtgactgttcgcgctgca 416
Db 864 TTATCCACCTGACCCAGCACAGTTAACAGACACATACAAAGATATTTATGCTTCA 923
QY 417 ggtgaagcaggattggtcctcaaggcaggttgacgtgtgtaatgacacagcgcagctcttt 476
Db 924 GCTTGGCAGGACATAGTTGCGAGGAGCTGTGCCCTGTCTTGTGCAACCTTAGCAATTAT 983
QY 477 gatttcaacattgtgcaatctgagattgggatttggatgaagccttgacagagagca 536
Db 984 AGGTTCTTACACATCCAGCTGAACTGGGAGACTACAGACCAGAACTCCATGGCGTGA 1043
QY 537 cttagcaaaaaataatacatatacc-----tcagcaagacgcactagagacaaatcgt 590
Db 1044 TTATGTTAGTATTTAAACTGGCCCGAATCAGACCAAGAACTTGAAGAGAGGTCTAT 1103
QY 591 ggaatttcaccataaacacattggacaacacacccagcagaaatcagatttccagctctcaga 650
Db 1104 GGAACATGCATAGTCATACAGGTCCATGACTCCAGCTCAGGCTGACTTGGAGTTTCTTGA 1163
QY 651 gattgcccctgctgtagatgatatggaatccggttgcccccgcagccagcagaggaagg 710
Db 1164 GAATGCCAAAAAGTTGTCTATGTATGGAGTTGATCTTCTATGAAGCAAGAGACTTGAAGG 1223
QY 711 caggaagatcaatctgcccgttgccaacacggaattctagttgttcagggtttcaactaa 770
Db 1224 AGTAGATATCATCCTAGGTGTCTGTCTAGTGGCTTCTGTTTCAAAAGATAAGCTGAG 1283
QY 771 gatcaatgccttcaactggtgccaaggtgcggaagctgagcttcaagaggaagcgctttct 830
Db 1284 AATTAAACGCTTCCCTTGGCCCCAAAGTGTCTGAAGATTCTTATAACGCTAGTAGCTTTT 1343
QY 831 catcaagctccggcagatgccaatagtcgctaccaggtataccttggaattcctgatggc 890
Db 1344 CATCAAGATTCCGGCTGGGAGACAGAGCAGTATGAAGTACCATCGGATTCAAACTTCC 1403
QY 891 cagtcgggatttctgcaagctctctggaaaatctgtgtgaacatcatgcctctcttag 950
Db 1404 CAGTTACCCAGCAGCTAAGAAATATGGAAGTCTGTGTAGACATCATCACAGTTTTTCAG 1463
QY 951 actttttgaagagcccaaacacaaagcccgctctctttagccgggggtcatcatt 1010
Db 1464 A---TTGACATCTACAGACACCATTCCTCCCAAGCAAAATTTCTTGCGCTAGGATCCAAAT 1520
QY 1011 tgggttcagtgtcggaactcagaagcag 1038
Db 1521 TCGATACAGTGGCCGAGCTCAAGCTCAG 1548

RESULT 12

AC022669 LOCUS AC022669 152053 bp DNA HTG 04-MAY-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-96B23 map 15, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
ACCESSION AC022669
VERSION AC022669.4 GI:9566291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 152053)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens chromosome 15, clone RP11-96B23

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 152053)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyns,S., Grant,G., Hags,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Sep 4, 2000 this sequence version replaced gi:7229803.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: 96_B_23

Center clone name: L4938

----- Summary Statistics

Sequencing vector: M13; M77815; 93% of reads

Sequencing vector: Plasmid; n/a; %0.1% of reads

7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of
reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136451 bases at least Q40

Consensus quality: 143612 bases at least Q30

Consensus quality: 146365 bases at least Q20

Insert size: 158000; agarose-fp

Insert size: 148853; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 33 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 784: contig of 784 bp in length

* 885 884: gap of 100 bp

* 885 1956: contig of 1072 bp in length

* 1957 2056: gap of 100 bp

* 2057 3113: contig of 1057 bp in length

* 3114 3213: gap of 100 bp

* 3214 4817: contig of 1604 bp in length

* 4818 4917: gap of 100 bp

* 4918 6582: contig of 1665 bp in length

* 6583 6682: gap of 100 bp

* 6683 8059: contig of 1377 bp in length

DEFINITION Human structural protein 4.1 mRNA, complete cds.
ACCESSION M14993
VERSION M14993.1 GI:182075
KEYWORDS Human reticulocyte, cDNA to mRNA, clones lambda-HE 4.1-8, lambda-HE
SOURCE 4.1-6, and lambda-HE 4.1-A.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2867)
AUTHORS Conboy, J., Kan, Y.W., Shohet, S.B., and Mohandas, N.
TITLE Molecular cloning of protein 4.1, a major structural element of the
human erythrocyte membrane skeleton
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 9512-9516 (1986)
MEDLINE 87092279
REFERENCE 2 (sites)
AUTHORS Conboy, J., Marchesi, S., Kim, R., Agre, P.P.A., Kan, Y.W. and
Mohandas, N.
TITLE Molecular analysis of insertion/deletion mutations in protein 4.1
in elliptocytosis; Determination of Molecular genetic origins of
rearrangements
JOURNAL J. Clin. Invest. 86, 524-530 (1990)
MEDLINE 90347048
COMMENT [2] sites: insertions and duplications causing elliptocytosis.
FEATURES
source Location/Qualifiers
1..2867
/organism="Homo sapiens"
/db_xref="taxon:9606"
799..2565
/note="4.1 protein"
/protein_id="AAA35795.1"
/db_xref="GI:182076"
/translation="MHCKVSLDDTVYECVVERHAKGQDLKRVCEHLNLEEDYFGL
AIVNATSKWLDALSKETIKQVGPWNFTNVPYPPDPAQLTEDITRYVLCQLRQ
DIVAGRLPCSPALALGYSYIQSELGDYDPELHGVYVDFKLPAPNOTKELEKVE
LHKSYSRMTQAQDLERENAKLSMYGVLDLHAKDLEGVDYIILGVSSGLLYVKDL
RINFPPKVLKIKSKSFYIKRPGEDQETISGIFKLPYSRAAKLWKVCVEHHT
KFAAASVDSAPSRKTPAPATQGVAGVLDASAKTKVVPKAKKTVKAEVKEDE
DPEAAPEPTAPMKRERLDGENIYIRHSLMLEDLDKSOEIKKHASISELKKNF
MESVPERPSPMDKRLTSHSPFTLNINGOIPTEGPPLVKTOFTVTSNANVAKSEI
PTKDVPIVHETKTIYENAGTVKGGISEFRIEKRIVTGDADIDHQLVQALKEAK
EQLPDMSVTKVNVQETEIADE"
variation 2016..2017
/note="duplication insertion of nucleotides 2217-2585
causing elliptocytosis [2]"
variation 2017..2256
/note="deletion causing elliptocytosis [2]"
BASE COUNT 932 a 620 c 657 g 658 t
ORIGIN 218 bp upstream of pvuII site.
Query Match 6.1%; Score 192.4; DB 9; Length 2867;
Best Local Similarity 52.8%; Pred. No. 1.8e-32;
Matches 490; Conservative 0; Mismatches 426; Indels 12; Gaps 3;
Qy 117 cgtgtccatcaaatccagatgctgatgacaccagaggcattgaagttcccaaaag 176
Db 798 CATGCACCTGCAAGGTTCTTGTGATGACACAGTTTATGAATGTTGTGGAGAAC 857
Qy 177 agctcctgggaagtgctgctggatgcggttgcaaccacccctcctggaagtgga 236
Db 858 TGCTAAGGACCAAGATTGCTTAACAGAGTATGTGAGCATCTCAATCTTTTGAAGAAGA 917
Qy 237 ctatttggcctcaggttctctcgcacaaaagatcacggttgctggtcctcctaaa 296
Db 918 CTAATTTGCTAGCCATTGGGATAACGCAACCTCTAAGACATGGCTGGATCCGCCAA 977
Qy 237 acccattgaaacagattagaagccaaagcacgctgtgttgtaagtttgttgtaaat 356
Db 978 AGAATAAAGACAGGTCGTGG---TGTCCTTGGAAATTATACATTAAATGTAAGTT 1034

Qy 357 ctttcgcgctgacacacacacaaactccaagaagaactcacaaggtacctgttgcgcgtcga 416
Db 1035 TTATCCACTTACCAGCAGCATAGTTAACAGACATATAAGATATATTATGTCTTCA 1094
Qy 417 ggtgaagcagactggctcaaggcagggttgctgctgtaataacacacagcgcagctctct 476
Db 1095 GCTTCGGCAGGACATAGTTGCGAGGAGCTGCGCTGCTCTTTCCTTGGCACTTATAT 1154
Qy 477 gatttcacacattgtgcaatctgagattggggatttttgatgaagccttgagacagagca 536
Db 1155 AGTTTCTTACACCATCCAGTCTGAAGTGGAGACTACGACCCAGAACTCCATGGCGTGA 1214
Qy 537 cttagcaaaaaataatacatatcc-----tcagcaagcagcactagagacacaaatcgt 590
Db 1215 TTATGTTAGTGTATTTTAACTTGCCCGCCGAATCAGACCAAGGAAGTGAAGGTCAT 1274
Qy 591 ggaatttcacacataaccacattggcaaacaccacagcagaatcagatttccagctcctaga 650
Db 1275 GGAAGTGCATAGTACAGGTCCATGCTCCAGCTCAGCTCAGCTGAGCTTGGAGTTCTTGA 1334
Qy 651 gattcccgctcggttagatgattccggttgccacccggcccaaggagcagaggaagg 710
Db 1335 GAATGCCAAAAAGTTGTCTATGATGGAGTTGATCTTCATAAAGCAAGGACTTGAAGG 1394
Qy 711 caggaagatcaatctggcgttgccaaacacgggaattcttagtttccaggtttcactaa 770
Db 1395 AGTAGATATCATCTAGTGTCTGCTCTAGTGGCTTTCGGTTTACAAAGATAAGCTGAG 1454
Qy 771 gatcaatgccttcaactggccagggtgcggaagctgcgagctcagctcgaagaggaagcgtttct 830
Db 1455 AATTAAACCGCTTCCCTTGGCCCAAGTGTCTGAAGATTCTTTATTAACGTAGTAGCTTTT 1514
Qy 831 catcaagctccgcgagatgccaatagctgctaccaggaatccttggaaattcctgatggc 890
Db 1515 CATCAAGATTCCGCTGGAGAGCAAGAGCAGTATGAAGTACCATCGGATTCAAACTTCC 1574
Qy 891 cagtcgggattcttcaagtcctcttggaataatctggtgaacatcatgcctcttttag 950
Db 1575 CAGTTACCCGACGAGTAAAGAAATATGGAAGTCTGTGTAGAACATCACACGTTTTTCAG 1634
Qy 951 acttttgaagagcaccacacaaagcccgccctctttagccgggggtcattcatt 1010
Db 1635 A---TTGACATCTACAGACACCATTTCCCAAAAGCAAAATTTCTTGGCTAGGATCCAAAT 1691
Qy 1011 tcggttcagtgctgcgactcagaagcag 1038
Db 1692 TCGATACAGTCCGCGACTCAAGCTCAG 1719
RESULT 14
HUMEMP41
LOCUS HUMEMP41 3043 bp mRNA PRI 03-MAR-1994
DEFINITION Homo sapiens erythroid membrane protein 4.1 mRNA, complete cds.
ACCESSION M61733
VERSION M61733.1 GI:182081
KEYWORDS erythroid membrane protein 4.1.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3043)
AUTHORS Conboy, J.G., Chan, J.Y.C., Chasins, J.A., Kan, Y.W. and Mohandas, N.
TITLE Tissue- and development-specific alternative RNA splicing regulates
expression of multiple isoforms of Erythroid membrane protein 4.1
J. Biol. Chem. 266, 8273-8280 (1991)
MEDLINE 91217063
FEATURES
source Location/Qualifiers
1..3043
/organism="Homo sapiens"
/db_xref="taxon:9606"
816..2741
/codon_start=1
/product="erythroid membrane protein 4.1"

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 12:00:11 ; Search time 283.82 Seconds
(without alignments)
9469.786 Million cell updates/sec

Title: US-09-555-342A-1_COPY_49_3183

Perfect score: 3135

Sequence: 1 atgggagaatagacagag.....acaaagagtctctgtgtat 3135

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3135	100.0	3442	20	AAH79183 Human chondrocyte-
2	3131.8	99.9	4687	21	AAA08582 Human cytoskeleton
3	1431.2	45.7	3094	21	AAQ98992 Human pancreatic c
4	223.6	7.1	241	21	AAA42150 Human secreted exp
5	169.6	5.4	3620	21	AAA08581 Human cytoskeleton
6	151.6	4.8	544	22	AAH98023 Murine 7-transmemb
7	144	4.6	3984	18	AAT58627 Protein tyrosine p
8	144	4.6	3984	20	AAH78463 Human PTPH1 CDNA
9	143.2	4.6	2872	22	AAH08887 Human protein tyro
10	133.6	4.3	3166	22	AAH17787 Human CDNA sequenc
11	95.8	3.1	659	22	AAH33845 Human colon cancer

12	85	2.7	665	22	AAH07867 Human CDNA clone (
13	85	2.7	1821	22	AAH15014 Human CDNA sequenc
14	76	2.4	410	21	AAH69420 Human secreted pro
15	69.4	2.2	1156	22	AAH99833 Human protein enco
16	63.8	2.0	1626	22	AAH59140 Human polynucleoti
17	63.8	2.0	2556	22	AAH60926 Human polynucleoti
18	62	2.0	1860	22	AAH99965 Human colon carcin
19	62	2.0	5923	22	AAH02052 DNA encoding molec
20	62	2.0	5923	22	AAH02052 DNA encoding molec
21	59.2	1.9	398	21	AAH30288 Human colon cancer
22	58.4	1.9	2355	22	AAH99674 Human protein enco
23	56.8	1.8	4080	19	AAH34368 Human protein tyro
24	56.8	1.8	4080	20	AAH34368 Human protein tyro
25	56.2	1.8	365	14	AAQ39923 Expressed Sequence
26	56.2	1.8	365	14	AAQ39923 Expressed Sequence
27	54.4	1.7	1778	22	AAH60338 Human brain Expres
28	54	1.7	550	22	AAH06244 Human CDNA clone (
29	52.8	1.7	556	22	AAH99189 Human protein enco
30	50.2	1.6	3115	21	AAH98113 Human colon cancer
31	48.2	1.5	454	22	AAH15794 Probe #5727 for ge
32	48.2	1.5	454	22	AAH15794 Probe #5727 for ge
33	48.2	1.5	1730	22	AAH62819 Human CDNA SEQ ID
34	48.2	1.5	1816	22	AAH58552 Human polynucleoti
35	47.4	1.5	2400	17	AAH42859 Choline oxidase ge
36	47.2	1.5	160	22	AAH25012 Probe #14945 for g
37	47.2	1.5	160	22	AAH25012 Probe #14945 for g
38	46.6	1.5	1696	22	AAH16459 Human CDNA sequenc
39	45.8	1.5	2400	18	AAH75000 Choline oxidase ge
40	44.6	1.4	347	21	AAH02360 Human secreted pro
41	44.6	1.4	625	22	AAH07308 Human CDNA clone (
42	44.4	1.4	342	21	AAH08796 Fusarium venenatum
43	44.2	1.4	10732	21	AAH10594 Gene encoding a su
44	44	1.4	463	15	AAQ68827 hgt4 clone showing
45	42	1.3	4610	21	AAH27896 Human GEF containi

ALIGNMENTS

RESULT 1

AAH79183
ID AAX79183 standard; DNA; 3442 BP.
XX
AC AAX79183;
XX
DT 17-AUG-1999 (first entry)
XX
DE Human chondrocyte-derived gene CDEP.
XX
KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;
KW drug; ss.
XX
OS Homo sapiens.
XX
PN WO9928458-Al.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-JP05348.
XX
PR 27-NOV-1997; 97JP-0342060.
XX
(CHUS) CHUGAI SEIYAKU KK.
XX
PI Kato Y, Kawamoto T, Koyano Y;
XX
DR WPI; 1999-371117/31.
XX
PT P-PSDB; AAY07482.
XX
PT Protein CDEP expressed in differentiated chondrocytes, and gene
XX
encoding it

PS Claim 5; Fig 1; 59pp; Japanese.

XX This sequence represents the coding region for a protein (CDRP) expressed
CC in differentiated human foetal chondrocytes, which contains an ezrin-like
CC domain, a DBI homology (DH) domain and a pleckstrin homology (PH) domain.
CC The nucleic acid or protein can be used in the investigation and
CC treatment of cancers and arthritic diseases (including chronic rheumatoid
CC arthritis), or for screening of candidate anticancer drugs.

XX
SQ Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;

Query Watch
Best Local Similarity 100.0%; Score 3135; DB 20; Length 3442;
Matches 3135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	atggggaaatagacagagggccgacccacgagatcacgactggggcccccggaaaattcg	60
DB	49	atgggaaatagacagagggccgacccacgagatcacgactggggcccccggaaaattcg	108
QY	61	ggatcagtcacttgaaacgtgacagaaagccgcccccaacaccttcaggaataactctg	120
DB	109	ggatcagtcacttgaaacgtgacagaaagccgcccccaacaccttcaggaataactctg	168
QY	121	tcacataaataccagatgctggatgacacccagagggcatttgaagtccacaaaagct	180
DB	169	tcacataaataccagatgctggatgacacccagagggcatttgaagtccacaaaagct	228
QY	181	cttggaaggtgctgtagtcagtttgcacacccactcaacctctgaaagtgcactat	240
DB	229	cttggaaggtgctgtagtcagtttgcacacccactcaacctctgaaagtgcactat	288
QY	241	tttggctcgagtttctctgatcacaaaaagatcacaggttggtggtatctctaaaccc	300
DB	289	tttggctcgagtttctctgatcacaaaaagatcacaggttggtggtatctctaaaccc	348
QY	301	attgtgaacagattagaaggccaaagcacgttgttgaagtgttggtgaaattcttt	360
DB	349	attgtgaacagattagaaggccaaagcacgttgttgaagtgttggtgaaattcttt	408
QY	361	ccgctgaccacacacaaactcaagaagaactcaaaagtacctgttcgcgtgcagggtg	420
DB	409	ccgctgaccacacacaaactcaagaagaactcaaaagtacctgttcgcgtgcagggtg	468
QY	421	aagcaggacttgctcaaggcagttgacgtgtaatgaccacagccagctctcttgatt	480
DB	469	aagcaggacttgctcaaggcagttgacgtgtaatgaccacagccagctctcttgatt	528
QY	481	tcacacattgtgcaatctgagattggggatttttgatgaagccttgacagagagcactta	540
DB	529	tcacacattgtgcaatctgagattggggatttttgatgaagccttgacagagagcactta	588
QY	541	gcaaaaaataatacatacctcagaagacgcgaactagagagacaaaaatctggaatttcac	600
DB	589	gcaaaaaataatacatacctcagaagacgcgaactagagagacaaaaatctggaatttcac	648
QY	601	cataaccacattggcaaaacaccagacagaatcagatttccagctcctagagattccctt	660
DB	649	cataaccacattggcaaaacaccagacagaatcagatttccagctcctagagattccctt	708
QY	661	cggtcagagatgtatggaatccggttgaccgcgccaagagacaggaagggacgaagatc	720
DB	709	cggtcagagatgtatggaatccggttgaccgcgccaagagacaggaagggacgaagatc	768
QY	721	aatctggcgttgcaaacacgggaattctagtttccagggtttccactaagatcaatgcc	780
DB	769	aatctggcgttgcaaacacgggaattctagtttccagggtttccactaagatcaatgcc	828
QY	781	ttaactgggccaaggtgcgggaagctgagcttcaagagaagcgcgtttctcatcaagctc	840
DB	829	ttaactgggccaaggtgcgggaagctgagcttcaagagaagcgcgtttctcatcaagctc	888
QY	841	cggtcagatgcgaatagtcgtaccaggataaccttggaattctctgatggccagtcgggat	900

DB	889	cggtcagatgcgaatagtcgtaccaggataaccttggaattctctgatggccagtcggat	948
QY	901	tcttgcagtcctctctggaaaatctgtgtgaacatcatcgcttcttttagacttttgaa	960
DB	949	tcttgcagtcctctctggaaaatctgtgtgaacatcatcgcttcttttagacttttgaa	1008
QY	961	gagcccaaaccaagcccaagcccgctcctcttttagcgggggtcatcattcgggtcagt	1020
DB	1009	gagcccaaaccaagcccaagcccgctcctcttttagcgggggtcatcattcgggtcagt	1068
QY	1021	ggtcggactcagaagcaggttctcgactatgtttaagaaggaggacataagaaggtgcag	1080
DB	1069	ggtcggactcagaagcaggttctcgactatgtttaagaaggaggacataagaaggtgcag	1128
QY	1081	tittgaagaagaacacagcaagatttctatccggagccttcttcacagcctacagaa	1140
DB	1129	tittgaagaagaacacagcaagatttctatccggagccttcttcacagcctacagaa	1188
QY	1141	ctgaattctggaaagtctgtagcagtcctcagcagaccagccttaccatttggagaaggt	1200
DB	1189	ctgaattctggaaagtctgtagcagtcctcagcagaccagccttaccatttggagaaggt	1248
QY	1201	gocgaattctccagggggccagagctccggcgaggaaagaaacccgaaggtttccgcggg	1260
DB	1249	gocgaattctccagggggccagagctccggcgaggaaagaaacccgaaggtttccgcggg	1308
QY	1261	gagccggggtgcaccccgagcctgcgcgagagagaagcccgcggttaacaagaaggcg	1320
DB	1309	gagccggggtgcaccccgagcctgcgcgagagagaagcccgcggttaacaagaaggcg	1368
QY	1321	gacggagccgctcgcgcacggaggaagagagaggtcttaaggatagagaccag	1380
DB	1369	gacggagccgctcgcgcacggaggaagagagaggtcttaaggatagagaccag	1428
QY	1381	cagagttaacctcagcccccgagcccaagccagcaggtcctcgtgactggcagtcctacct	1440
DB	1429	cagagttaacctcagcccccgagcccaagccagcaggtcctcgtgactggcagtcctacct	1488
QY	1441	tcgagctctctgtaactcgcagggggagtgccctgcccaagtcaccttgctctcc	1500
DB	1489	tcgagctctctgtaactcgcagggggagtgccctgcccaagtcaccttgctctcc	1548
QY	1501	aacctgagcccgacacccagcaggtcctccttgatcagcccgctgctggaatgaccag	1560
DB	1549	aacctgagcccgacacccagcaggtcctccttgatcagcccgctgctggaatgaccag	1608
QY	1561	gocgtcccccgagcgacgatgaggtgagggccggagagaagagattcccaactgataa	1620
DB	1609	gocgtcccccgagcgacgatgaggtgagggccggagagaagagattcccaactgataa	1668
QY	1621	gcgtactctatagctaagaagtgtctaccacgcaggaacacatctcgaagatctcgaa	1680
DB	1669	gcgtactctatagctaagaagtgtctaccacgcaggaacacatctcgaagatctcgaa	1728
QY	1681	gttatacctctggttttcagagcacagtgcgcaagagagacccatgcccgaagcactg	1740
DB	1729	gttatacctctggttttcagagcacagtgcgcaagagagacccatgcccgaagcactg	1788
QY	1741	aaaagtctcatattcccgaaatttgaacctttgcacaaatttccataactatttctcaag	1800
DB	1789	aaaagtctcatattcccgaaatttgaacctttgcacaaatttccataactatttctcaag	1848
QY	1801	gaaattgagcaacgacttgcctctgtggaggccgctcaaatgcccacatcaagattac	1860
DB	1849	gaaattgagcaacgacttgcctctgtggaggccgctcaaatgcccacatcaagattac	1908
QY	1861	caagaatcggcgatgctcgtgaagaacatttcaggggatgagcacctggcggtcac	1920
DB	1909	caagaatcggcgatgctcgtgaagaacatttcaggggatgagcacctggcggtcac	1968
QY	1921	ctgtggaaagcacagcgagccttggaggcccttggaatggaaatcaagagctccccggcg	1980

Qy	3061	atggaagtatccgcagtgccaccagctctgctcgacccacgctgttagccacaaa	3120
Db	3109	atggaagtatccgcagtgccaccagctctgctcgacccacgctgttagccacaaa	3168
Qy	3121	gagtcctctgtgtat	3135
Db	3169	gagtcctctgtgtat	3183
RESULT 2			
AAA08582			
ID	AAA08582 standard; DNA; 4687 BP.		
XX			
AC	AAA08582;		
XX			
XX	20-JUL-2000 (first entry)		
DT			
XX	Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.		
DE			
XX			
XX	Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;		
KW	autoimmunity; inflammatory, vesicle trafficking; neurological;		
KW	cardiovascular; cell motility; reproductive; muscle disorder; ss.		
KW			
XX			
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
FH	90..3227		
FT	CDS		
FT	/*tag=a		
FT	/product= CYSKP-2		
XX			
XX	W0200017355-A2.		
XX			
XX	30-MAR-2000.		
XX			
PF	17-SEP-1999; 99WO-US21565.		
XX			
XX	18-SEP-1998; 98US-0172226.		
PR	27-APR-1999; 99US-0131321.		
PR			
XX			
PA	(INCY-) INCYTE PHARM INC.		
XX			
XX	Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;		
PI	Guegler KJ, Patterson C, Azimzai Y, Baughn MR;		
PI			
XX			
DR	WPI; 2000-283582/24.		
DR	P-PSDB; AAY91947.		
DR			
XX			
PT	Human cytoskeleton associated proteins, used to treat cell		
PT	proliferative, autoimmune/inflammatory, vesicle trafficking,		
PT	neurological, cell motility, reproductive and muscle disorders		
XX			
PS	Claim 9; Page 101-102; 113pp; English.		
XX			
CC	AAA08581-96 encode human cytoskeleton associated proteins 1 to 16		
CC	(CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat		
CC	and diagnose cancer and cell proliferative, autoimmune/inflammatory,		
CC	vesicle trafficking, neurological, cardiovascular, cell motility,		
CC	reproductive and muscle disorders. Pharmaceutical compositions		
CC	containing CYSKP-2 to CYSKP-16 can be used to treat or prevent disorder		
CC	associated with decreased expression or activity of CYSKP (claimed), f		
CC	example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,		
CC	cancers, autoimmune/antiinflammatory disorders such as allergies, anem		
CC	asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,		
CC	diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma		
CC	and trauma. CYSKP antagonists can be used to treat or prevent a		
CC	disorder associated with increased expression or activity of CYSKP		
CC	(claimed).		
XX			
SQ	Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;		

99.98: score 3131.8; DB 21; Length 4687;

Best Local Similarity 99.9%; Pred. No. 0;				
Matches 3133; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	1	atggagaaatagacagagccgaccacccagcagatcacgactggggcccgaaaaattcg	60	
Db	90	atgggaaatagacagagccgaccacccagcagatcacgactggggcccgaaaaattcg	149	
QY	61	gggacagtcacttgaaacgtggacagaaagccgcccccaacacctcagagaaactcgtg	120	
Db	150	gggacagtcacttgaaacgtggacagaaagccgcccccaacacctcagagaaactcgtg	209	
QY	121	tcacataaatccagatgctgatgacacccaggagcatttgaagtccacaagaagct	180	
Db	210	tcacataaatccagatgctgatgacacccaggagcatttgaagtccacaagaagct	269	
QY	181	cctgggaaggtgctgctgatgacagtttcaaccacctcaacctcgtggaagtgactat	240	
Db	270	cctgggaaggtgctgctgatgacagtttcaaccacctcaacctcgtggaagtgactat	329	
QY	241	tttggcctcagtttctctgatcacaaaaagatcaacggtggtgctggtatctctaaaccc	300	
Db	330	tttggcctcagtttctctgatcacaaaaagatcaacggtggtgctggtatctctaaaccc	389	
QY	301	attgtgaacagattagaagccaaagcacgtttgtttaagtttgtgtaaatctctt	360	
Db	390	attgtgaacagattagaagccaaagcacgtttgtttaagtttgtgtaaatctctt	449	
QY	361	cgcctgacacacacaaactccaagaagaaactcacaaagtacctgttcgcgtcaggtg	420	
Db	450	cgcctgacacacacaaactccaagaagaaactcacaaagtacctgttcgcgtcaggtg	509	
QY	421	aagcagacttgctcaagcgaggttgacgtgtaatagcacacagcgacgtctcttgatt	480	
Db	510	aagcagacttgctcaagcgaggttgacgtgtaatagcacacagcgacgtctcttgatt	569	
QY	481	tcacacattgtgcaatctctgagattggggattttgtatgaagccttgacagagacactta	540	
Db	570	tcacacattgtgcaatctctgagattggggattttgtatgaagccttgacagagacactta	629	
QY	541	gcaaaaaataatatacactcagcaagcagcactagagacaaaaactcgtgaatttcac	600	
Db	630	gcaaaaaataatatacactcagcaagcagcactagagacaaaaactcgtgaatttcac	689	
QY	601	cataaccatttgcaaacacacagcagatcagatttccagctcctcagagattgccgt	660	
Db	690	cataaccatttgcaaacacacagcagatcagatttccagctcctcagagattgccgt	749	
QY	661	cggctagagatgtatggaatccggttgccacccggccaaaggacagggaaagcagaatcc	720	
Db	750	cggctagagatgtatggaatccggttgccacccggccaaaggacagggaaagcagaatcc	809	
QY	721	aacttgccgttgcaaacacggaattctagtgtttcagggtttcactaagatcaatgcc	780	
Db	810	aacttgccgttgcaaacacggaattctagtgtttcagggtttcactaagatcaatgcc	869	
QY	781	ttaactggccaaagtgcggaagctgagcttcaagagaaagcgctttctcacaagctc	840	
Db	870	ttaactggccaaagtgcggaagctgagcttcaagagaaagcgctttctcacaagctc	929	
QY	841	cggccagatgccaatagtgcgtaccaggaataccttggaaattcctgatggccagtcgggat	900	
Db	930	cggccagatgccaatagtgcgtaccaggaataccttggaaattcctgatggccagtcgggat	989	
QY	901	tcttgaaatcctcttgaaaaatcgtgttgaaatcatgcctctcttagaacttttgaa	960	
Db	990	tcttgaaatcctcttgaaaaatcgtgttgaaatcatgcctctcttagaacttttgaa	1049	
QY	961	gagcccaaaccaagcccaagccgctctcttagccgggggtcatcatttcggttcagt	1020	
Db	1050	gagcccaaaccaagcccaagccgctctcttagccgggggtcatcatttcggttcagt	1109	
QY	1021	ggtcgactcagaagcaggttctcgactatgttaagaaggaggacataagaaggtgcag	1080	

Db	1110	ggtcgactcagaagcaggttctcgactatgtttaagaaggaggacataagaaggtgcag	1169	
QY	1081	tttgaagaagacacagcaagattcattctatccgagagcttgccttcacagcctacagaa	1140	
Db	1170	tttgaagaagacacagcaagattcattctatccgagagcttgccttcacagcctacagaa	1229	
QY	1141	ctgaattcgaagagtcgtgagcagtcctcagcagagcaccagccttacatttggagaagt	1200	
Db	1230	ctgaattcgaagagtcgtgagcagtcctcagcagagcaccagccttacatttggagaagt	1289	
QY	1201	gccgaatctccggggccagagctgcggcgaggaagaaacccgaaggtttccgcggg	1260	
Db	1290	gccgaatctccggggccagagctgcggcgaggaagaaacccgaaggtttccgcggg	1349	
QY	1261	gagccgggttcacccagccctgcgcgaggaagaccccgcggttaacaagcagcg	1320	
Db	1350	gagccgggttcacccagccctgcgcgaggaagaccccgcggttaacaagcagcg	1409	
QY	1321	gagcagccgctcgcgcgaggaagagagaggttcgtttaaggatagagcccg	1380	
Db	1410	gagcagccgctcgcgcgaggaagagagaggttcgtttaaggatagagcccg	1469	
QY	1381	cagagtaaacctcaagcccgagcccaagcacaggtcctcactgagtcagtcacatt	1440	
Db	1470	cagagtaaacctcaagcccgagcccaagcacaggtcctcactgagtcagtcacatt	1529	
QY	1441	tccgagctctgtgaactcgcagggggagtgggccctgcgaacgtgaccttgcctccc	1500	
Db	1530	tccgagctctgtgtgaactcgcagggggagtgggccctgcgaacgtgaccttgcctccc	1589	
QY	1501	aacctgagcccgacacacagcagcctcctctgtcagcccgctcgtgtaatgaccag	1560	
Db	1590	aacctgagcccgacacacagcagcctcctctgtcagcccgctcgtgtaatgaccag	1649	
QY	1561	gctgcccccgagcgacgatgagatgagggccggaggaagagattcccaactgataa	1620	
Db	1650	gctgcccccgagcgacgatgagatgagggccggaggaagagattcccaactgataa	1709	
QY	1621	gcgtactctatagctaaggaagtgtctaccacgcagcgaacatactgaagatctcgaa	1680	
Db	1710	gcgtactctatagctaaggaagtgtctaccacgcagcgaacatactgaagatctcgaa	1769	
QY	1681	gttatacctctggttttcagagcacagtgcgaagagagcgccatgcgcgaagcactg	1740	
Db	1770	gttatacctctggttttcagagcacagtgcgaagagagcgccatgcgcgaagcactg	1829	
QY	1741	aaaagtctcaattccggaatttgaacctttgcacaaatttcaactaatttttccaag	1800	
Db	1830	aaaagtctcaattccggaatttgaacctttgcacaaatttcaactaatttttccaag	1889	
QY	1801	gaaattgagcaacgacttgcctgtgggaagccgctcaaatgcccacatcagagattac	1860	
Db	1890	gaaattgagcaacgacttgcctgtgggaagccgctcaaatgcccacatcagagattac	1949	
QY	1861	caaaagaatcgcgatgtcatgtctgaagaacatttcaggggcatgaagcacctcgggtcac	1920	
Db	1950	caaaagaatcgcgatgtcatgtctgaagaacatttcaggggcatgaagcacctcgggtcac	2009	
QY	1921	ctgtggaagcacagcgagccttggagccctggagaaatggaaatcaagagctcccgcg	1980	
Db	2010	ctgtggaagcacagcgagccttggagccctggagaaatggaaatcaagagctcccgcg	2069	
QY	1981	ctggagaactctcagagacattgagctgcagaaggtgtgttaacctacgcctcaacacc	2040	
Db	2070	ctggagaactctcagagacattgagctgcagaaggtgtgttaacctacgcctcaacacc	2129	
QY	2041	ttcctcctcgccactgcacccggtcattgcactacaagcaggtccttgagcggtgtgc	2100	
Db	2130	ttcctcctcgccactgcacccggtcattgcactacaagcaggtccttgagcggtgtgc	2189	
QY	2101	aaacaccccccgcagccagccgacgttcaggagactgcgagccgctttggcagagatc	2160	
Db	2190	aaacaccccccgcagccagccgacgttcaggagactgcgagccgctttggcagagatc	2249	

QY	2161	acggaagatggttgccacagctccacggtacgatgataagaatggagaatttccagaagctg	2220
Db	2250	acggaagatggttgccacagctccacggtacgatgataagaatggagaatttccagaagctg	2309
QY	2221	cacgaactcaagaagaatttgatggcattgacaactcttggttcgcggaaaggagttc	2280
Db	2310	cacgaactcaagaagaatttgatggcattgacaactcttggttcgcggaaaggagttc	2369
QY	2281	atccgtctgggcagcctcagcaagctctcggggaaggggctccagcagcgcatgtttcttc	2340
Db	2370	atccgtctgggcagcctcagcaagctctcggggaaggggctccagcagcgcatgtttcttc	2429
QY	2341	ctgttcaacgacgctctgctatacacagcgcggggctgacgcgctccaatcagtttaa	2400
Db	2430	ctgttcaacgacgctctgctatacacagcgcggggctgacgcgctccaatcagtttaa	2489
QY	2401	gtccacggcagctcccgcctctatggcatgacgattgaggagagcgaagcagtggggg	2460
Db	2490	gtccacggcagctcccgcctctatggcatgacgattgaggagagcgaagcagtggggg	2549
QY	2461	gtccccactgctaacctccggccagcgacgagtcacatcgtggcccgacgttct	2520
Db	2550	gtccccactgctaacctccggccagcgacgagtcacatcgtggcccgacgttct	2609
QY	2521	cgtccgagatggagaagtgggttgaggacatccagatggccattgacctggcgagagaag	2580
Db	2610	cgtccgagatggagaagtgggttgaggacatccagatggccattgacctggcgagagaag	2669
QY	2581	agcagcagcccccttgagttctctgcccagcagcccccttgacacacagtcacctgat	2640
Db	2670	agcagcagcccccttgagttctctgcccagcagcccccttgacacacagtcacctgat	2729
QY	2641	gaagccaccgcggtgaccagagtcagagagatgacctgagcgctcgccacatcgctg	2700
Db	2730	gaagccaccgcggtgaccagagtcagagagatgacctgagcgctcgccacatcgctg	2789
QY	2701	gagccagcccccgccagccgcccgaacacatggtgcacgtgtgctggcaccgcacaccc	2760
Db	2790	gagccagcccccgccagccgcccgaacacatggtgcacgtgtgctggcaccgcacaccc	2849
QY	2761	agcgctccatggtggacttcagctcagctgcagtgagaatcagttgctgaaacctgctg	2820
Db	2850	agcgctccatggtggacttcagctcagctgcagtgagaatcagttgctgaaacctgctg	2909
QY	2821	aggaaattcaaaaaagcaacgggtggcagaagctgtgggtgtgttcacaaactttcgc	2880
Db	2910	aggaaattcaaaaaagcaacgggtggcagaagctgtgggtgtgttcacaaactttcgc	2969
QY	2881	gtttttctcaaatcacacagacaatcatcccttcgacgctgcctctgctgcgcgc	2940
Db	2970	gtttttctcaaatcacacagacaatcatcccttcgacgctgcctctgctgcgcgc	3029
QY	2941	tactcgtcaccatccctctgagtcgcgagaacatccagaagactacgtgttcaagctg	3000
Db	3030	tactcgtcaccatccctctgagtcgcgagaacatccagaagactacgtgttcaagctg	3089
QY	3001	cacttcaagtcacagctctacttacttcaggcggaagcagagtcacacgtttcgaaagtgg	3060
Db	3090	cacttcaagtcacagctctacttacttcaggcggaagcagagtcacacgtttcgaaagtgg	3149
QY	3061	atggaagtgatccgagtgccaccagctctgctcgcgaccccccagctgtttgagccacaaa	3120
Db	3150	atggaagtgatccgagtgccaccagctctgctcgcgaccccccagctgtttgagtcacaaa	3209
QY	3121	gagctctctgtgtat 3135	
Db	3210	gagctctctgtgtat 3224	

RESULT 3

AAC98992

ID AAC98992 standard; cDNA; 3094 BP.

[illegible]

AAC98992;

09-MAR-2001 (first entry)

Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.

Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; cytostatic; neuroprotective; neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological; antiinflammatory; candidate; gene therapy; chromosome mapping; linkage analysis; tissue identification; tissue typing; forensic; neural; immune system; muscular; reproductive; gastrointestinal; pulmonary; cardiovascular; renal; proliferative; ss.

Homo sapiens.

WO200055320-A1.

21-SEP-2000.

08-MAR-2000: 2000WO-US05989.

12-MAR-1999: 99US-0124270.

(HUMA-) HUMAN GENOME SCI. INC.

Rosen CA, Ruben SM:

WPI: 2000-579444/54.

P-FSDB; AABJ4227.

NEW NUCLEIC ACID
treating, or amel

Claim 1; Page 664-665; 1379pp; English.

PC98773 to AAC59231 encode the human pancreatic cancer associated proteins, called pancreatic cancer antigens, given in AAB54008 to AAB54466. The human pancreatic cancer antigens have cytostatic, neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive, gynaecological, cardiant and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer.

antigens and antagonists to the antigens can be screened for. The pancreatic cancer antigens polynucleotides can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99322 to AAC99240 and AAC85467 represent sequences used in the exemplification of the present invention.

Sequence 3094 BP: 736 A: 849 C: 755 G: 749 T: 5 other:

Query Match	45.7%	Score 1431.2;	DB 21;	Length 3094;
Best Local Similarity	99.8%	Pred. No. 0;		

Qy 1692 gtggtttcagagcacagtgagcaaaaggagcgcctgccggaagcactgaaaagtctcat 1751

Qy 1752 attccggaattttgaacatttgcacaaatttcacataactaattttctcaagaaaatttgagca 1811

Db 122 atcccgaaattgaaacattgacaaatttcataactaattttctcaaggaaattgagca 181
QY 1812 acgacttgccttgggaagccgctcaaatgccaaatcagagattaccaaagaatcgg 1871
Db 182 acgacttgccttgggaagccgctcaaatgccaaatcagagattaccaaagaatcgg 241
QY 1872 cgtatgctatgctgaagaacattcagggtcgaagcattggaagcctggcgccttgggaagca 1931
Db 242 cgtatgctatgctgaagaacattcagggtcgaagcattggaagcctggcgccttgggaagca 301
QY 1932 cagcgaagccttggagccctgagaaatggaatcaagagctcccgcgctggagaaactt 1991
Db 302 cagcgaagccttggagccctgagaaatggaatcaagagctcccgcgctggagaaactt 361
QY 1992 ctgcaagacttggagctcagaggtgtgttacctaccgctcaaacaccttctcctcg 2051
Db 362 ctgcaagacttggagctcagaggtgtgttacctaccgctcaaacaccttctcctcg 421
QY 2052 gccactgcaccgctcatgcactacaagcaggtcctggagcgtgtgcaaacaccaccc 2111
Db 422 gccactgcaccgctcatgcactacaagcaggtcctggagcgtgtgcaaacaccaccc 481
QY 2112 gccagccacgcgaacttcaggactccgagccgcttggcagagatcacggagatggt 2171
Db 482 gccagccacgcgaacttcaggactccgagccgcttggcagagatcacggagatggt 541
QY 2172 gccagactccagctacgatgataagatggaagatttccagaagctgcagaaactcaa 2231
Db 542 gccagactccagctacgatgataagatggaagatttccagaagctgcagaaactcaa 601
QY 2232 gaaagatttgattggcattgacaatctgtgttcggaagagggttccatccgtctgg 2291
Db 602 gaaagatttgattggcattgacaatctgtgttcggaagagggttccatccgtctgg 661
QY 2292 cagcctcagcaagctctcgggaaggggtcccgagcagcagatgttctcgtttcaacga 2351
Db 662 cagcctcagcaagctctcgggaaggggtcccgagcagcagatgttctcgtttcaacga 721
QY 2352 cgtcctgctatacacagccggggtgacgctccatcagtttaagtcacagggca 2411
Db 722 cgtcctgctatacacagccggggtgacgctccatcagtttaagtcacagggca 781
QY 2412 gctccgctctatggcatgacgattgaggagagcgaagcagatgggggtgcccaactg 2471
Db 782 gctccgctctatggcatgacgattgaggagagcgaagcagatgggggtgcccaactg 841
QY 2472 cctgacccctcgggcccagcagtcacatcctgctggccgagttctcgttccagat 2531
Db 842 cctgacccctcgggcccagcagtcacatcctgctggccgagttctcgttccagat 901
QY 2532 ggagaagtggttgagacatccagatggcattgacctggcgggagaaagcagagccc 2591
Db 902 ggagaagtggttgagacatccagatggcattgacctggcgggagaaagcagagccc 961
QY 2592 cgccctgagttcctggcagcagcccccctgacaaagtcacctgatgaagccaccgc 2651
Db 962 cgccctgagttcctggcagcagcccccctgacaaagtcacctgatgaagccaccgc 1021
QY 2652 ggctgaccagagtcagagatgacctgagcctcgcacatcgcctgagcgcagggc 2711
Db 1022 ggctgaccagagtcagagatgacctgagcctcgcacatcgcctgagcgcagggc 1080
QY 2712 ccgcaccgcggcaacacaaatggtgacgtgtgctggcaccgcaaacaccagcgtccat 2771
Db 1081 ccgcaccgcggcaacacaaatggtgacgtktgtctggcaccgcaaacaccagcgtccat 1140
QY 2772 ggtggacttcagcatcgagtgagaaatcagttgtgtgaaacctgctgaggaattcaa 2831
Db 1141 ggtggacttcagcatcgagtgagaaatcagttgtgtgaaacctgctgaggaattcaa 1200
QY 2832 aaacagcaacgggtgcagaagctgtgggtgtgtgttcacaaactctgctgttctcta 2891
|||||

Db 1201 aaacagcaacgggtgcaagaagctgtgggtgtgttcacaaactctgctgttctcta 1260
QY 2892 caaatcacaccaggaacaatcatcccttgcagcctgctctgctcggtactcgcctcac 2951
Db 1261 caaatcacaccaggaacaatcatcccttgcagcctgctctgctcggtactcgcctcac 1320
QY 2952 catccctctgagtcgcgagaacatccagaagactacgtgttcaagctgcaactcaagtc 3011
Db 1321 catccctctgagtcgcgagaacatccagaagactacgtgttcaagctgcaactcaagtc 1380
QY 3012 ccagctctactactcagggcggaagcaggtacacagttcgaaggtggatggaagtgt 3071
Db 1381 ccagctctactactcagggcggaagcaggtacacagttcgaaggtggatggaagtgt 1440
QY 3072 ccgagtgccaccagctctgctcgcgaccccgctgttgagccacaaagactctctgt 3131
Db 1441 ccgagtgccaccagctctgctcgcgaccccgctgttgagccacaaagactctctgt 1500
QY 3132 gtat 3135
Db 1501 gtat 1504
RESULT 4
AAA42150/c
ID AAA42150 standard; cDNA; 241 BP.
XX
AC AAA42150;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:890.
XX
KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; ESR; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vacciner; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021990-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24205.
XX
PR 15-OCT-1998; 98US-0104435.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
DR WPI: 2000-317937/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (ESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 354; 618pp; English.
XX
CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (ESTs), isolated from human, mouse, xenopus and rat

tissue sources. The sESTs can have a range of activities depending on the tissues they were isolated from. The activities include:

- chemotactic; proliferative; immunomodulatory; haematopoietic;
- chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
- cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
- antiasthmatic; vulnerrary; antiparasitic; osteopathic; neuroprotective;
- neotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
- anticonvulsant; and antidepressant. The sESTs can be used for gene therapy and in vaccines. The sESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the sESTs. Proteins encoded by the sESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA43420 to AAA43425 represent linker variants which are given in the exemplification of the present invention.

Sequence 241 BP: 59 A; 58 C; 58 G; 66 T; 0 other:

Sequence 241 BP; 59 A; 58 C; 58 G; 66 T; 0 other;

[illegible]

RESULT	5	
AAA08581		
ID	AAA08581	standard; DNA; 3620 BP.
XX	XX	
XX	AC	AAA08581;
XX	XX	
DT	19-JUL-2000	(first entry)
XX	XX	
DE	XX	Human cytoskeleton associated protein 1 (CYSKP-1) coding sequence
XX	XX	
KW	KW	Cytoskeleton associated protein; CYSKP-1; cancer; proliferative;
KW	KW	autoimmunity; inflammatory, vesicle trafficking; neurological;
KW	KW	cardiovascular; cell motility; reproductive; muscle disorder; ss.

PR 18-SEP-1998; 98US-0172226.
PR 27-APR-1999; 99US-0131321.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
XX
XX WPI; 2000-283582/24.
DR P-PSDB; AAY91946.
DR
XX
XX Human cytoskeleton associated proteins, used to treat cell
PT proliferative, autoimmune/inflammatory, vesicle trafficking,
PT neurological, cell motility, reproductive and muscle disorders
XX
XX Example 5; Page 100-101; 113pp; English.
PS
XX
XX AA08581-96 encode human cytoskeleton associated proteins 1 to 16
CC (CYSPK-1 to CYSPK-16) respectively. The sequences can be used to treat
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,
CC vesicle trafficking, neurological, cardiovascular, cell motility,
CC reproductive and muscle disorders. Pharmaceutical compositions
CC containing CYSPK-2 to CYSPK-16 can be used to treat or prevent disorders
CC associated with decreased expression or activity of CYSPK (claimed), for
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis
CC cancers, autoimmune/antiinflammatory disorders such as allergies, anem
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
CC and trauma. CYSPK antagonists can be used to treat or prevent a
CC disorder associated with increased expression or activity of CYSPK
CC (claimed).
CC

Sequence 3620 BP: 1156 A: 773 C: 906 G: 785 T: 0 other:

[illegible]

Db 1206 aaaccacaggggttatccgagcacaaagctgattccagttctttagaaaaatgcaag 1265
 QY 661 cggtagagatgatgaatccggttgaccgcccgaagcagaggaagcagcaagatc 720
 Db 1266 agccttccatgatggtgtgacctacatcatgccaagagactcagaaggtggaacatc 1325
 QY 721 aatctggccgttgccacacggaattctagtgtttcagggtttcactaagatcaatgcc 780
 Db 1326 aagctggcggtgtgctaatgacttctcatattacaagacagactgcgaatcaatcgt 1385
 QY 781 ttaactggccgaaggtgcggaagctgagcttcaagaggaagcgtttctcacaagctc 840
 Db 1386 ttgctggccgaatcttcaaaattctctataaagcgaactctctacataaagtc 1445
 QY 841 cgccagatccaatagctgtaaccagataccttgaaattctctgattgcccagtcggat 900
 Db 1446 agaccggcagagctggaacggttgagagaccattggaattcgaactgcgaaccaccgg 1505
 QY 901 tctgcaagctctctggaaaactctgttgaacatcatgctcttcttagactttttgaa 960
 Db 1506 gcagcgaagagactggaagtgctggtgagcatcatttctacaggtgtt--- 1562
 QY 961 gagcccaaaccaagcccaagccgctctctttagccggggtcactcattcgttcagt 1020
 Db 1563 tctccagagcagcccaaaagcgaagttctgacttggtggtccaaaattcgctatagt 1622
 QY 1021 ggtcggactcagaagcag 1038
 Db 1623 ggcgcaccaacagcag 1640

RESULT 6
 ID AAH98023 standard; DNA; 544 BP.
 AC AAH98023;
 XX
 XX
 DT 10-OCT-2001 (first entry)
 XX Murine 7-transmembrane G-protein coupled receptor coding sequence #267.
 XX Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;
 KW 7-transmembrane G-protein coupled protein receptor; ds.
 XX Mus sp.
 XX WO200160999-A1.
 XX PD 23-AUG-2001.
 XX 14-FEB-2001; 2001WO-US04700.
 XX 14-FEB-2000; 2000US-0182377.
 XX (IMCL-) IMCLONE SYSTEMS INC.
 PA (UYP-) UNIV PRINCETON.
 XX Lemischka IR, Witte L, Pereira DS;
 XX WPI; 2001-522596/57.
 DR DNA Sequences encoding 7-transmembrane G-protein coupled protein
 PT receptors characteristic of hematopoietic stem cells, useful for
 PT treating leukemia -
 XX Claim 1; Page 122; 176pp; English.
 PS The present invention relates to murine coding sequences for
 CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The
 CC present sequence is one such murine 7TM-GPCR coding sequence. The present
 CC sequence was derived from stromal stem cells. The present sequence
 CC and its corresponding protein are useful in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate 7TM-GPCR expression.

CC 7TM-GPCRs identify specific signalling molecules, to activate an
 CC effector-signalling cascade that triggers an intracellular response and
 CC eventually a biological effect.
 XX Sequence 544 BP; 143 A; 131 C; 150 G; 120 T; 0 other;
 SQ

Query Match 4.8%; Score 151.6; DB 22; Length 544;
 Best Local Similarity 56.3%; Pred. No. 1.3e-30;
 Matches 304; Conservative 0; Mismatches 234; Indels 2; Gaps 1;
 QY 2231 agaaagatttgattgacattgtgacaaattctgtgttccgggaagggagttcattcgtctg 2290
 Db 4 agcgagacgtggtggtgtagaaacctcattgtctctgggaggtttatcgtcgagg 63
 QY 2291 gaagctcagcaagctctcgggaaggggtccagagcagcatgttcttctcttcaacg 2350
 Db 64 gctgctgcaagactcaccagaagggcgtgcaagagaggtttttcttctctcag 123
 QY 2351 acgtctctatatacagcgcggggtgacgctcccaatcagtttaaaagtcacagggc 2410
 Db 124 ataigtgtgtatatacaagcaaaagtgtcacagagccagtcatttccggatcgtggt 183
 QY 2411 agctccgctctatgacatgacgattgagagagcgaagagtgagggtggtgcccact 2470
 Db 184 tcttccactcgtggtgtagaaagtgagaatgaatggtctgttctctcatt 243
 QY 2471 gctgacctcggggccagcggcagtcctcagtcagtcggtggtcagtcgttcgagaga 2530
 Db 244 gcttccactctatgacgtcagaacaattgtgtgtagcagcagcactcggctagaaa 303
 QY 2531 tgagaagtggttgaggacatccagatggcattgacgtggtggaagagcagcagcc 2590
 Db 304 agaaaagtgtgtagcagcagcgtgaatgcagcaatccaagcagcagcagcagtcgtgact 363
 QY 2591 cggccctgagttcctggccagcagccccctgacaaagtcctcctgtatgaagccacog 2650
 Db 364 caccctcagtgctggtggagggccgtgtatactgtacctgtacctgtatctctgatgaag 423
 QY 2651 cggctaccagagtcagagtagcctgagcgcctcgcgcacatcgtgagcagcagc 2710
 Db 424 tctctggaagatcagaagatggtcagaggaacccggggtccctcggaggggaac--ag 481
 QY 2711 cccgcacccggcgaacacaatggtgcacgtgtgctggcaccgcaacacacagcgtctcca 2770
 Db 482 ccagcaccggggcaatacaacaatgcatgtgtgtgtaccgtataatacaagtgtctcca 541

RESULT 7
 ID AAT58627 standard; cDNA; 3984 BP.
 AC AAT58627;
 XX
 XX
 DT 22-MAY-1997 (first entry)
 XX Protein tyrosine phosphatase cDNA.
 DE
 DE Protein tyrosine phosphatase; PTPH1; focal adhesion;
 KW protein tyrosine kinase; malignancy; cancer; gene therapy;
 KW retrovirus; vector; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 24..2765
 FT /*tag= a
 XX US5595911-A.
 XX 21-JAN-1997.
 PD
 XX 14-MAR-1990; 90US-0494036.
 PF

[illegible]

Qy	657	cctgcggtcgtatgatgaataccggtttgaccocggccaaggacaggaaggcacgaa	716
Db	656	gcggacctcgacttctatgtagtagaacgtgcacagtggtagggattcgacaaattaga	715
Qy	717	gatacaatcgcggttcaccaacacggaattctagttcttcagggtttcactaatgacaa	776
Db	716	cctaattgattggaattgcttcgcggcgcttgctgtgtaccgaaaaatacatattgcacaag	775
Qy	777	tgccttcaactggcgccaaggtgcggaagctgagcttccaagaggaagcgctttctcatcaa	836
Db	776	ttctatccttggtgaacattctcaaaatttcttccaaaaggaaaaagtcttcataca	835
Qy	837	gctccgccagatgccaaatagtcggtaccaggataaccttggaattctctgatggccagtcg	896
Db	836	tcagcgacagaaaacagcgtgaatcagggaacatatattgtgcttcaacatgctgaatta	895
Qy	897	ggatttctgaagtccttcggaaaactgtgttgaaacatcagctctcttt	948
Db	896	ccgattctgcaaaactitggaaaatcctgttgtgagcaccatacgtctcttt	947
RESULT	8		
AAAX78463	ID	AAAX78463 standard; cDNA; 3984 BP.	
XX	AC	AAAX78463;	
XX	DT	07-SEP-1999 (first entry)	
XX	DE	Human PTPH1 cDNA.	
KW	PPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;		
KW	localisation; treatment; overexpression; oncogenic; cell transformation;		
KW	prevention; phosphotyrosine; disease; malignant; ss.		
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
FT	CDS	24..2765	
FT		/tag= a	
FT		/product= "PTPH1"	
FT		111..1094	
FT	misc_feature	/tag= b	
FT		/note= "Region of homology to the N-terminal domain	
FT		of band 4.1, ezrin and talin. This region is	
FT		known to be important for localisation to focal	
FT		adhesions"	
PN	US5863781-A.		
XX	26-JAN-1999.		
XX	04-DEC-1996;	96US-0759536.	
PR	01-MAR-1991;	91US-0663579.	
PR	14-MAR-1990;	90US-0494036.	
PR	16-AUG-1993;	93US-0107420.	
PR	04-DEC-1996;	96US-0759536.	
PA	(COLD-) COLD SPRING HARBOR LAB.		
XX	Tonks NK;		
XX	WPI; 1999-131308/11.		
DR	P-PSDB; AAY25156.		
XX	Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells		
PT	Disclosure; Fig 1A-B; 12pp; English.		
XX	This sequence encodes a novel protein tyrosine phosphatase, PTPH1		
CC	isolated from HeLa cells. The protein of the invention appears to		
XX			

localise to focal adhesions and is therefore potentially useful in the treatment of cancer. Overexpression of pPrl1 can be used to counter the effects of oncogenic protein tyrosine kinases such as those of transforming viruses and for interfering with or reversing cell transformation. This would provide a means of preventing or reversing abnormally high levels of phosphotyrosine associated with any disease or condition such as preventing or reversing malignancy associated with the activity of a protein tyrosine kinase.

Sequence 3984 BP; 1086 A; 954 C; 930 G; 1014 T; 0 other;

Query Match 4.6%; Score 144; DB 20; Length 3984;
Best Local Similarity 50.8%; Pred. NO. 4e-28;
Matches 423; Conservative 0; Mismatches 400; Indels 9; Gaps 3;

[illegible]

RESULT	9
AAS08887	
ID	AAS088
XX	
AC	AAS088

ID AAS08887 standard: cDNA: 2872 BP.

AAS08887:

DT 26-SEP-2001 (first entry)

Human protein tyrosine phosphatase (PTPase) DNA #2.

KW Protein tyrosine phosphatase; PTPase; human; nervous system; dementia;
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; autism;
 KW Tourette's syndrome; obsessive compulsive disorder; schizophrenia; mania;
 KW cardiovascular disorder; atherosclerosis; myocarditis; Addison's disease;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; vitiligo;
 KW ankylosing spondylitis; Sjogren's syndrome; asthma; atopie dermatitis;
 KW diabetes mellitus; inflammatory disorder; pancreatitis; sarcoidosis; ss;
 KW allogeneic transplant rejection; multiple myeloma; Burkitt's lymphoma;
 KW leukaemia; cancer; neoplasm.

OS Homo sapiens.

Key	Location/Qualifiers
-----	---------------------

FT	CDS	230..1339
----	-----	-----------

```
FT      /*tag= a
FT      /product= "Human PTPase #2"
```

AA WO200153530-A1.

26-JUL-2001.

17-JAN-2001:

AA
PR 18-JAN-2000: 2000US-0176306.

PA (HUMA-) HUMAN GENOME SCI INC.

Shi Y, Ruben SM:

WPI; 2001-442265/47.

XX
XX
1994-1995

PT New human protein tyrosine phosphatase polypeptide useful for treatment of disorders including disorder of the nervous system -

PS Claim 1; Page 282-283; 290pp; English.

CC The sequence represents DNA encoding a human protein tyrosine phosphatase
CC (Ptpase) polypeptide. Ptpase proteins and nucleic acids are useful for
CC preventing, treating or ameliorating a medical condition. The medical
CC conditions are disorders of the central and peripheral nervous system
CC (e.g. Alzheimer's disease, Huntington's disease, Parkinson's disease,
CC Tourette's syndrome, obsessive compulsive disorder, schizophrenia, mania,
CC dementia and autism), cardiovascular disorders (e.g. atherosclerosis and
CC myocarditis), autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, ankylosing spondylitis, Addison's disease, Sjogren's syndrome,
CC vitiligo, asthma, atopic dermatitis and diabetes mellitus), inflammatory
CC disorders (e.g. pancreatitis, sarcoidosis and allogeneic transplant
CC rejection) and cancers and neoplasms (e.g. leukaemia, Burkitt's lymphoma
CC and multiple myeloma). The presence or absence of a mutation in the
CC nucleic acid encoding the protein allows for diagnosis of or
CC susceptibility to a pathological condition. The sequences are also useful
CC in screening for agonists and antagonists of ptpase activity.

Sequence 2872 BP: 840 A; 575 C; 601 G; 856 T; 0 other;

Query Match	Score	DB 22:	Length
Best Local Similarity	4.68;	143.2;	2872;
Matches 490: Conservative	50.3%;	Pred. No. 5.5e-28;	
Matches 490: Mismatches	0;	468;	
Indels	16;	Gaps	5

Qy 83 gacgaagccgcccccaacaccttcaggaaaactcgtgtccatcaaatccagatgctgg 142


```
Db 408 tggattcagcacaaagtagcacattggttgatggttacaaaagcatcaaaaagcaagtaa 467
Qy 317 gaagggcaaaagcagctgttgtaagtgttggtgaattcttccgctcgaccacacac 376
Db 468 aaattggttccacctattgtctcatctcogattgaagttttattctcagaaccaata 527
Qy 377 aactccaagaagaactcaaacagtagctgttcgctgcaggtgaagcaggaacttggctc 436
Db 528 acctcgtgaggagtaaacccggtattatttcttcagtttaaaacaagatatctca 587
Qy 437 aagggcaggttgacgtgtaataacaccagcgcagctctcttgatttccacacattgtgcaat 496
Db 588 gtggaataattagactgtccctttgatcacagcagtgcaattggcagcttataatctgcaag 647
Qy 497 ctgagattggggtttttagaagcccttgacag---agagcacttagcaaaaaataaat 553
Db 648 ctgaacttgggtactatgattctgtgagcatagtcctgaacttgtctcagagttcagat 707
Qy 554 acatacc---tcagcaagacgcactagaggacaaaactcgtggaatttcaccataaacaca 610
Db 708 tcgtgcattcagactgaagagatggaactggctatttttgagaatggaagaataca 767
Qy 611 ttggcaaacacacagcagaatcagatttccagctcctcagagattgcccgtcggtcagaga 670
Db 768 gaggtcaaacacacagcagcagctgaaaccaaattatctgataaagcacaatggctagaaa 827
Qy 671 tgtatggaatccggttgcaccggccaagacagaggaaggaagcagcagatcaatctggccg 730
Db 828 tgtatggggtgatgtgcattggtgcaagctagagatggaatgactagtttgggac 887
Qy 731 ttgccacaacggaaattctagtgttcaggtttcactaagatcaaatgccttcaactggg 790
Db 888 taacacacacagagctcgtgttttgaaggagataccaaaattggctattttttggc 947
Qy 791 ccaaggtgpgaaagctgagcttcaagaggaagcgtttctcatcaagctccggccagatg 850
Db 948 cgaagataaccagatggattttcaagaagaataaataaaccttgggtttagaagatg 1007
Qy 851 ccaat----agtgcgtaccagataccttggaaattcctgagtcgagtcgggatttct 904
Db 1008 atgacagggcaaaagacaggaacacacatttgcctttagactcggatcccaaaagcat 1067
Qy 905 gcaagtcctcttgaaaaatcgtgtgtaacatcatgcctctttagacttttgaagagc 964
Db 1068 gcaacatttatgaaatgctgtgagcatcatcgtcttctccgcttcgagcccg 1127
Qy 965 ccaaaacaaagcccaagc---ccgtcctctttagccgggggtcattcattcggttcagtg 1021
Db 1128 tccaaagagttctcatcgatcagcaggatttattcogactagatcagcagatttagatagt 1187
Qy 1022 gtcggactcagaagcaggttctcgtactatgtttaaagaaggaagacataaagaagtgcat 1081
Db 1188 ggaacacagatgctcagaccacaaaaccaaataaagaagaagatcaa-----catcct 1241
Qy 1082 ttgaaggaagcagacgaagattcattctatccggagcct 1121
Db 1242 ttgaagaaggccagcaacagatatcttagacgaactct 1281
```

RESULT 11

AAH33845

ID AAH33845 standard; cDNA; 659 BP.

XX

AC AAH33845;

XX

XX 03-SEP-2001 (first entry)

DE

XX Human colon cancer antigen encoding cDNA SEQ ID NO:901.

XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; ss.

XX

OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR P-PSDB; AAG74414.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 2820; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 659 BP; 170 A; 147 C; 172 G; 163 T; 7 other;

Query Match 3.1%; Score 95.8; DB 22; Length 659;
Best Local Similarity 53.6%; Pred. No. 1.3e-15;
Matches 199; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy 137 tcttgatgacacccagggagcatttgaagttccacaaagagcctcctgggaagtgctgc 196
Db 58 tctcagcgggagcgaagtgcgctggacctgcgcaaacatgccaaagccaggtattgt 117
Qy 197 tggatgcagtttgcaaccacacctcaacctcctgggaagtgactattttgcccagatttc 256
Db 118 tigtatcagattgtgtaccactggaccttgggaacagatacttcttggcctccagttcc 177
Qy 257 ctgatcacaaaaagatcacgggtggtggtggtctctctctctctctctctctctctctct 316
Db 178 tctgactctgccaggttgcgactggtgctgctgctgctgctgctgctgctgctgctgctg 237
Qy 317 gaaggccaaagcacgttctgttgaagtttgcgtgaaattcttccgctgcgaccacac 376
Db 238 aaattggacctgcttattgtcttaccttgcagtttgaatactattcttctcagaaccaaca 297
Qy 377 aactccaagaagaactcacaaggtacctgttcgcgtgcaggtggaagcaggttggtctc 436
Db 298 acctcgtgaggagtttacaaggtacctgttcttttacaactcagcagcatgacattctt 357
Qy 437 aaggcaggttgacgtgttaataacacacagcagcagctctcttatttccacacattgtgcaat 496
Db 358 ctggaaaaattgaatgccttatgaaacagctgtggaattagctctctctctctctcaag 417

CC increase or decrease storage capabilities. AAC69390 to AAC69398 and
CC AAB38118 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 410 BP; 100 A; 93 C; 105 G; 112 T; 0 other;

Query Match 2.4%; Score 76; DB 21; Length 410;
Best Local Similarity 94.0%; Pred. No. 2.1e-10;
Matches 79; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1013 ggttcaggtgctgactcagaacaggtctcactatgttaagaaggagcataaga 1072

Db 53 gtttcaggtgctgactcagaacaggtctcactatgttaagaaggagcataaga 112

QY 1073 aggtgcagttgaaaggaacaca 1096

Db 113 aggtgcagttgaaaggaagtaga 136

RESULT 15

AAH99833

ID AAH99833 standard; cDNA; 1156 BP.

XX

AC AAH99833;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human protein encoding cDNA sequence SEQ ID NO:568.

XX

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder;; ss.

XX Homo sapiens.

OS

XX WO200153455-A2.

PN

XX 26-JUL-2001.

PD

XX 22-DEC-2000; 2000WO-US35017.

XX

XX 23-DEC-1999; 99US-0471275.

PR

PR 21-JAN-2000; 2000US-0488725.

PR

PR 25-APR-2000; 2000US-0552317.

XX

XX (HYSE-) HYSEQ INC.

XX

XX Tang YT, Liu C, Drmanac RT;

PI

XX WPI; 2001-457603/49.

DR

XX P-PSDB; AAM25892.

XX

XX Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX

XX Claim 1; Page 674; 1217pp; English.

XX

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to

CC AAM25963. The proteins can have activities based on the tissues and

CC cells they are expressed in, such as: antiinflammatory; antirheumatic;

CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX SQ Sequence 1156 BP; 381 A; 202 C; 255 G; 318 T; 0 other;

Query Match 2.2%; Score 69.4; DB 22; Length 1156;
Best Local Similarity 50.1%; Pred. No. 2.1e-08;
Matches 262; Conservative 0; Mismatches 246; Indels 15; Gaps 3;

QY 608 acattggacaacacacagcagaatcagattccagctccttagagattgcccgctgctag 667

Db 381 acagaggtcaacacacagcagcgtgaaaccaattatctgaataaagcaaatggctag 440

QY 668 agatgtatgaatccggttcaccccgccgaagacagcagggcagcagcaagaataatcgg 727

Db 441 aaatgtatgggttgatgatcgtgtgtcgaagctagatgagtggaatgactatagttgg 500

QY 728 ccgttgccaacacggaattctagtgttcagggttcactaagatcaatgcctcaact 787

Db 501 gactaacacacacagcagcgtcctgttttgaaggagataccaaaattggctatttttt 560

QY 788 gggcccaaggtgcggaagctgagcttcaagaggaagcgtcttctcatcaagctccgccaag 847

Db 561 ggcgaagataacacagattgattttaagaagaataataaacttggtgttagaag 620

QY 848 atgccaat-----agtcgtaccagataaccttggaattcctctgattggcagctgggatt 901

Db 621 atgatgatcaggggcaagaacaggaacatacatttgtcttagactggatccatccaaag 680

QY 902 tctgcaagtccttctgaaaaatctgtgtgaacatcgcctcttcttagactttttgaag 961

Db 581 catgcaaacatttatggaaatgtgtgtggagcaatcgtttcttcgccttcgagggcc 740

QY 962 agcccaaaccaagcccaagc---ccgtcctctttagccggggtcatcatttctgttca 1018

Db 741 ccgtccaaaagagtctctcatcgatcagcagatttattcactaggatcacgattagata 800

QY 1019 gtggtcggactcagaacagcaggttctcagactatgttaagaaggagagacataaagaagtgc 1078

Db 801 gtggaaaacacagatcatcagaccacaaaacccaataagaagaagatcaaa-----cat 854

QY 1079 agtttgaaggagcagcaagattcttctatcccgagcct 1121

Db 855 cctttgaaggagcccgcaagcaacgatatctcttagacgaactct 897

Search completed: December 6, 2001, 12:01:00

Job time: 11372 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 14:32:44 ; Search time 102.73 Seconds
(without alignments)
6911.404 Million cell updates/sec

Title: US-09-555-342A-1_COPY_49_3183
Perfect score: 3135
Sequence: 1 atgggagaaatagacagag.....acaaagagtcttctgtat 3135

Scoring table:
IDENTITY_NUC
Gapex 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.8	1.8	4080	2 US-08-446-345-35	Sequence 35, Appl
2	53.8	1.7	7218	1 US-08-232-463-14	Sequence 14, Appl
3	47.4	1.5	2400	4 US-08-930-001-1	Sequence 1, Appl
4	44	1.4	463	1 US-07-752-101A-67	Sequence 67, Appl
5	42.2	1.3	4411529	4 US-09-103-840A-1	Sequence 1, Appl
6	41.2	1.3	4403765	4 US-09-103-840A-2	Sequence 2, Appl
7	40.8	1.3	1598	1 US-08-211-682-24	Sequence 24, Appl
8	40.8	1.3	4776	2 US-08-852-401-1	Sequence 1, Appl
9	40.4	1.3	4791	4 US-08-949-155-49	Sequence 49, Appl
10	40	1.3	2277	1 US-08-676-967-5	Sequence 5, Appl
11	40	1.3	2277	1 US-09-098-487-5	Sequence 5, Appl
12	40	1.3	2277	2 US-09-098-487-5	Sequence 8, Appl
13	38.6	1.2	2574	2 US-08-677-734A-8	Sequence 51, Appl
14	38.4	1.2	4267	4 US-08-949-155-51	Sequence 19, Appl
15	38.2	1.2	1053	6 5352575-6	Sequence 19, Appl
16	37.6	1.2	2619	4 US-08-467-822-19	Sequence 19, Appl
17	37.6	1.2	2619	4 US-08-432-697-19	Sequence 31, Appl
18	37.6	1.2	2619	4 US-08-466-248-19	Sequence 31, Appl
19	37.6	1.2	2634	1 US-08-196-218-31	Sequence 31, Appl
20	37.6	1.2	2634	1 US-08-681-953-31	Sequence 2, Appl
21	37.6	1.2	3468	1 US-07-951-715A-2	Sequence 4, Appl
22	37.6	1.2	3468	1 US-07-951-715A-4	Sequence 2, Appl
23	37.6	1.2	3468	2 US-08-459-448A-2	Sequence 2, Appl
24	37.6	1.2	3468	2 US-08-459-448A-2	Sequence 2, Appl
25	37.6	1.2	3468	3 US-08-459-595A-2	Sequence 2, Appl
26	37.6	1.2	3468	3 US-08-459-595A-4	Sequence 2, Appl
27	37.6	1.2	3468	3 US-08-459-504B-2	Sequence 2, Appl

28	37.6	1.2	3468	3 US-08-459-504B-4	Sequence 4, Appl
29	37.6	1.2	3468	3 US-08-459-444-2	Sequence 2, Appl
30	37.6	1.2	3468	3 US-08-459-444-4	Sequence 4, Appl
31	37.6	1.2	3468	3 US-09-053-549-3	Sequence 3, Appl
32	37.6	1.2	3468	3 US-09-053-549-5	Sequence 5, Appl
33	37.6	1.2	3475	4 US-09-657-481A-10	Sequence 10, Appl
34	37.6	1.2	3476	3 US-08-630-916A-47	Sequence 47, Appl
35	37.2	1.2	2745	1 US-08-363-255-1	Sequence 1, Appl
36	37.2	1.2	2745	1 US-08-363-255-13	Sequence 13, Appl
37	37.2	1.2	44377	2 US-08-804-227C-7	Sequence 7, Appl
38	37.2	1.2	44377	2 US-08-804-198-1	Sequence 3, Appl
39	36.8	1.2	2376	1 US-07-912-952-3	Sequence 3, Appl
40	36.6	1.2	1273	4 US-09-319-892-3	Sequence 3, Appl
41	36.6	1.2	1288	1 US-08-440-856A-9	Sequence 9, Appl
c 42	36.4	1.2	2824	2 US-09-010-928B-3	Sequence 3, Appl
43	36.4	1.2	3509	1 US-08-322-962-1	Sequence 1, Appl
44	36.4	1.2	3509	3 US-08-450-653-1	Sequence 1, Appl
c 45	36	1.1	2214	3 US-08-864-038A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-446-345-35
; Sequence 35, Application US/08446345
; Patent No. 5831009
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-446-345-35

Query Match 1.8%; Score 56.8; DB 2; Length 4080;
Best Local Similarity 46.6%; Pred. No. 2.5e-05;

Matches 218; Conservative 0; Mismatches 247; Indels 3; Gaps 1;

```
QY 130 atccagatgctgagatgacacccaggagcatttgaagttccacaaagagctcctgggaag 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 411 ATCCAAGTCTGTAATACAGAGTTGTGGAGTTTACCTGTCCGTGGAGACACTGGCCAG 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 190 gtgctgctgagtcagtttgcacccactcaactcgtggaagtgactattttggcctc 249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 471 GAAAGCCTCGAGCCCGTGGCCAGAGCTGGAGCTGGGGAGGTCACCTTACTTCAGCCTC 530
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 250 gagtttctgatacacaacaaagatacaogtgtggtggtgactctcctaaacccattgtgaaa 309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 531 TGGTACTACAAACAAGCAAGAAATACAGCGCGTGGGTAGATTGGAAAAACCTTTGAAGAAG 590
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 310 caqat---tagaaggccaagcagctgttggtaagttgtgtaaatcttccgcct 366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 591 CAGCTGGATAAATATGCAATTGGAACTTACCGTCTATTTTGGAGTGGTGTATGTGCCT 650
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 367 gaccacacaaactccaagaagaactcacaaaggtacctgttcgctgcaggtgaagcag 426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 651 TCAGTTTCTCAGCTGCAGCAGAGATTAACAGGTATCAGTATTATCTGCAACTGAAGAAA 710
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 427 gacttgcctcaaggcaggttgacgtgtaatagcacacagcagctctctctgattcaacac 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 711 GATATCTTGAAGGAAGTATTCCTGTGTACCTTAGAACAGCAATTCAGCTAGCAGGCTTA 770
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 487 attgtcaatctagattgggatttggatgaagccttgacagagacgacttagcaaaa 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 GCTGTTCAACGGATTTTGTGTGACTTTGTATGATCAGTAYGAATCCAGGACTTCTTCAGAAA 830
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 547 aataaaatacatactcagcagcagcactagagagacaaaatcgtggaa 594
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 831 TTTCCTTGTCTCTGTGGGATGTTTACAGATGAAGATAAGTATGGAA 878
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 2

```
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; TITLE OF INVENTION: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
```

```
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14
```

Query Match 1.7%; Score 53.8; DB 1; Length 7218;
Best Local Similarity 4.7%; Pred. No. 0.00021;
Matches 19; Conservative 222; Mismatches 164; Indels 0; Gaps 0;

```
QY 1010 ttcggttcagtgctcgagactcagaagcaggttctcgaactatgttaaagaggagacata 1069
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1437 TACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1378
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1070 agaagtgagtttgaagaagacacagcaagattcattctatccgagccttgctcac 1129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1377 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1318
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1130 agctacagaaacttcggaagtgcagcagctctcagcagagcaccagccttacct 1189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1317 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1258
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1190 ttgagaagtgccgaatctccagggggccagagctgcggcgaggaagaaacccgaagg 1249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1257 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1250 ttcccgccgggagccgggtcgacccagcctgcgcgagagagagcccgcggtga 1309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1197 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1310 acaagcagcgagcagcgccctcgcccgaggaagagagagagagagagagagag 1369
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1137 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1078
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1370 ataggccacagagataaacctcagcccgcccgagccaacagcacag 1414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1077 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1033
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 3

```
US-08-930-001-1
; Sequence 1, Application US/08930001A
; Patent No. 6281412
; GENERAL INFORMATION:
; APPLICANT: MURATA, No. 628141210
; TITLE OF INVENTION: METHOD FOR PRODUCING OSMOTOLERANT PLANTS
; FILE REFERENCE: 0230-118P
; CURRENT APPLICATION NUMBER: US/08/930,001A
; EARLIER FILING DATE: 1997-09-26
; EARLIER APPLICATION NUMBER: JAPAN 106819/1995
; EARLIER FILING DATE: 1995-03-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2400
; TYPE: DNA
; ORGANISM: Arthrobacter globiformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (361)..(2001)
; US-08-930-001-1
```

Query Match 1.5%; Score 47.4; DB 4; Length 2400;
Best Local Similarity 43.3%; Pred. No. 0.0063;
Matches 222; Conservative 0; Mismatches 291; Indels 0; Gaps 0;

Db 838877 CCGCGCGCGCGCGCGTTG 838859

RESULT 6
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007 00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 1.3%; Score 41.2; DB 4; Length 4403765;
Best Local Similarity 45.8%; Pred. No. 11;
Matches 142; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
QY 2478 cctccggggccagcggcagtcacatcgtggtggccgagttctcgttcgagatggagaa 2537
Db 1174714 CGTCACCGCGGAGCATCTCACCAGGTGTTCCGACATCGCGGTGATCGACATCGACGA 1174655
QY 2538 gtgggtgagacatcagatgcccattgacctggcggagaaagacagcagccccc 2597
Db 1174654 CGGGGTGGCGTTACCTCTGACACGAGCGTGCAGGAAATTCGCGAGCGGCGGACTA 1174595
QY 2598 tgaattctggcagcagcccccctgacaacagtcctcctgatgaagccaccggtgta 2657
Db 1174594 CCGGGCCTGCGGTGCGAGTGCATATGTCGTGCGACCGTGCAGGGGATCGCGCCTG 1174535
QY 2658 ccaggagtcagagatgacctgagcgcctgcgcacatgctggagcgcagggcccgca 2717
Db 1174534 GGATGTGTCCACCGGGAACCGCATCGCGCTGGCCACACCGGGTGACCATCGACCGGAT 1174475
QY 2718 ccggcgcaacacatggtgacatgtgctggaccgcaacaccagcgtctccatgtga 2777
Db 1174474 CTTGCGGAGCGCATACACTCTGGGCTACGCGCCGAGACCATCATCGCGGAGAGGG 1174415
QY 2778 cttcagcatc 2787
Db 1174414 AGTGACCATC 1174405

RESULT 7
US-08-211-682-24
; Sequence 24, Application US/08211682
; Patent No. 5670333
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: EXPRESSION OF POLYPEPTIDES IN E.COLI UNDER
; TITLE OF INVENTION: CONTROL OF THE E.COLI MDH-GENE PROMOTER
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,682
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1598 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 286..1269
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 228..233
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 205..210
US-08-211-682-24

Query Match 1.3%; Score 40.8; DB 1; Length 1598;
Best Local Similarity 45.2%; Pred. No. 0.3;
Matches 150; Conservative 0; Mismatches 182; Indels 0; Gaps 0;
QY 2305 cctccgggaaggggtccagcagcagtccttctcttcaacgacgtccgtctatc 2364
Db 415 CCCAGGCCATGAAGGCCCTGGAGGCGTGGTCATGGAGCTGGAGGACTGCGCCTTCCCC 474
QY 2365 acgagcgggggtgacgcccctccaatcagtttaaaagtccacgggagcgtcccgcctctat 2424
Db 475 CTGCTTGGCGGCTTCAGGCCACCGACGCCCAAGTGCCTTCAAGGACGCGGACTAC 534
QY 2425 ggcagcagcattgagagagcgaagacagagtgggggtgccccactgacctgacccctccg 2484
Db 535 GCCTCTGCTGTGGGGCGGCGCCCGCAAGCGGGGATGGAGCGCGGACCTTTTTCAG 594
QY 2485 ggcagcggcagtcacatcgtggcgcagttctcgttcgagatggagagtggtt 2544
Db 595 GTAACGGCAAGATCTCACCGAGAGCGGCGGCGCCTGGCCGAGGTGGCCCAAGAAGGAC 654
QY 2545 gaggcacatcagatggccattgacctggcgagagaagacagcagcagcccccctgagttc 2604
Db 655 GTCAAGGTGCTGTGTGGCAACCCCGCCCAACACACACCCCTCATCGCCTACAGAAC 714
QY 2605 ctggcagcagccccctgacacaaagtcccc 2636
Db 715 GCGCCGCGGCTCAACCCCGGAACCTTCACGCG 746

RESULT 8
US-08-852-401-1
; Sequence 1, Application US/08852401
; Patent No. 5976836
; GENERAL INFORMATION:
; APPLICANT: Weber, J. Mark
; APPLICANT: Hessler, Paul E.
; APPLICANT: Larsen, Peter E.
; APPLICANT: Luu, B. Minh
; TITLE OF INVENTION: Methods and Compositions for Enhancing
; TITLE OF INVENTION: Erythromycin Production
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:

COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Review #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/949,155
FILING DATE: Concurrently Herewith
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/027,338
FILING DATE: 11-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,094
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: 41,071
REFERENCE/DOCKET NUMBER: TAMK:177
TELEPHONE INFORMATION:
TELEPHONE: (512) 418-3000

```

, INFORMATION FOR SEQ ID NO: 49:
,
, SEQUENCE CHARACTERISTICS:
,     LENGTH: 4791 base pairs
,     TYPE: nucleic acid
,

```

US-08-949-155-49
TOPOLOGY: linear
;
; STRANDEDNESS: single

Query Match	1.3%	Score 40.4;	DB 4;	Length 4791;
Best Local Similarity	59.6%;	Pred. No. 0.66;		
Matches 68;	Conservative	0;	Mismatches. 46;	Indels 0;
				Gaps 0;

Qy	2521	2866	2925
	cggtccgagatgggagaagtgggttgaggacatccagatggccattgacctggcgagaag	ccgccaggctctggagaagagagactggagccttgaggggaaacgtccctcgaggaccag	

Qy	2581	agcagcagcccgcccctgaattctctgggccagcagccccctgacacaagtcc	2634
Db	2926	AGTCGAGCTCTCCCTTAAGCTGCTGCCGTGCCCGTCCCCACCCGGCACCCC	2979

RESULT 10
US-08-676-967-5

Sequence 5, Application US/086/696/
Patent No. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA

```

; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: 08/676 067

```

; FILING DATE: 10/1/2010
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 36,627


```

QY 2505 cgtggccgcagtcttcggtccagatggaagaagtgggttgaggacaatccagatggccat 2564
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 804 GAACGGCGCGCTGAAGCGCCCGCCCGCCCGCCCAAGAGCAGCAGCAGGAGGACAG 863
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2565 tgacctggcgagaagacagacagcccccgccctgagttcctggccagcagccccctga 2624
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 864 CGACTGGAGAGACGACAGCATCGACGGCGAGAGCTGGCCCGACGAGCCACACCAG 923
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2625 caacaagtccctgatgaagccaccgcg 2652
      ||||| ||||| ||||| |||||
Db 924 CACCGAGGAGCAGGAGGACAAGGCCGTG 951
      ||||| ||||| ||||| |||||

RESULT 13
US-08-677-734A-8
; Sequence 8, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2574 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-677-734A-8

Query Match 1.2%; Score 38.6; DB 2; Length 2574;
Best Local Similarity 44.1%; Pred. No. 1.5;
Matches 161; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY 2272 agggagttcatccgtctgtggcagctcagcaagctctcggggaagggtctccagacgcgc 2331
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1762 AGAGAAAATGTCAGCGTGTCTGCCTGGACATGCAGTCTCTGGAGCAGCGAGCGGCGGAGC 1821
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2332 atgttcttcctgttcaacgagctcctgtatatacagccgggggtgacggcctccaat 2391
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1822 ATCCGGAGCGGAGGAGCATGGTCAGCCACACAGCTATACAGATGACTGTATACAGCGG 1881
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2392 cagttaaagtccacggggcagctccgcctctatggcatgacattgagagagcgaagac 2451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
Db 3020 GCCTCTCTGCGCCAGGAGGCTGATAGAGGAGCATGAAGGAGGTGAAGGCGCTACCGC 3079
QY 1387 aaactcagcccccagcagcaagcacagctccctgactggcagtcctcactttccgag 1446
Db 3080 GAGGAGCTGAGCGCGCTGGCCCGCTGACCCAGGAGACGCGCGCGCTGTCCAAG 3139
QY 1447 ctgtctgtgaactcgagggggagtgggccctgccaacgtgaccttgtctcccaacctg 1506
Db 3140 GAGCTGAGCGCGCGCGCGCGCTGGCGCGGACATGAGGAGCGTGGCGCAACCGCTTG 3199
QY 1507 agccccgacaccaagcaggcctctccttgatcagccctgctgtaataaccaggcctgc 1566
Db 3200 GTGCTCTAGCGCAGCGAGGTGCACAAACATGTTGGCGCAGACCACCGAGGAGCTGCGGAGC 3259
QY 1567 ccccgagac 1574
Db 3260 CGCCTGGC 3267
```

RESULT 15

5352575-6

; Patent No. 5352575

; APPLICANT: PETROVSKIS, ERIK A.; POST, LEONARD E.; TIMMINS, JAMES G.

; TITLE OF INVENTION: PSEUDORABIES VIRUS PROTEIN

; NUMBER OF SEQUENCES: 12

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/513,282

; FILING DATE: 20-APR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 100,817

; FILING DATE: 29-JUN-1987

; APPLICATION NUMBER: 886,260

; FILING DATE: 16-JUL-1986

; APPLICATION NUMBER: 784,787

; FILING DATE: 04-OCT-1985

; APPLICATION NUMBER: 801,799

; FILING DATE: 26-NOV-1985

; APPLICATION NUMBER: 844,113

; FILING DATE: 26-MAR-1986

; SEQ ID NO: 6:

; LENGTH: 1053

; 5352575-6

Query Match

1.2%; Score 38.2; DB 6; Length 1053;

Best Local Similarity 54.0%; Pred. No. 1.2;

Matches 101; Conservative 0; Mismatches 83; Indels 3; Gaps 1;

```
QY 1215 gggccagagctgcgcgcgaggaacccgaaggtttccgcgcgggagccgggtcgca 1274
Db 564 gtggcagcagcctgcgtgctgtgaccccgccgagagccggtgttaccacgcgcgc 623
QY 1275 cccgagc---cctgcgcagagagagcccccgcgggttaacaaagcagcgagcgccgc 1331
Db 624 ccgcatcgagcagagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 683
QY 1332 ctgcgcgcagcagagagagaggtcgttaagatagagcccccagcagtaaac 1391
Db 684 cgagccccctcccgacagagagagagagagagagagagagagagagagagagagagag 743
QY 1392 tcagccc 1398
Db 744 ggtgccc 750
```

Search completed: December 6, 2001, 16:58:20
Job time: 30017 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 10:34:44 ; Search time 2912.37 Seconds
(without alignments)
11567.217 Million cell updates/sec

Title: US-09-555-342A-1_COPY_49_3183
Perfect score: 3135
Sequence: 1 atggggagaatagacagag.....acaaagagtctctgtgtat 3135

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rpd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	784.8	25.0	1122	10	BE746268
2	761.2	24.3	804	11	BG747792
3	757.2	24.2	865	10	BE260677
4	755.8	24.1	784	11	BG763918
5	753	24.0	930	10	AL566821
6	745	23.8	861	11	BG764061
7	726.2	23.2	901	11	BG475554
8	716	22.8	736	11	BF793662
9	705.4	22.5	898	10	BE745887
10	704.6	22.5	835	11	BG829192
11	700.2	22.3	902	11	BE910036
12	687.6	21.9	721	10	AU132546

13	679.6	21.7	942	11	BG323704
14	678.2	21.6	891	11	BG767698
15	673	21.5	883	11	BF206296
16	660.6	21.1	981	11	BG171514
17	660	21.1	680	11	BG825738
18	628	20.0	859	10	AL537488
19	625.4	19.9	741	11	BG122769
20	616.8	19.7	690	11	BG698789
21	608.4	19.4	968	11	BF206873
22	607.6	19.4	934	11	BF314265
23	606	19.3	621	11	BG770181
24	605.6	19.3	692	11	BG116225
25	599.4	19.1	890	11	BE907778
26	598.8	19.1	704	11	BG769615
27	590.4	18.8	932	11	BF686586
28	572.8	18.3	788	11	BG750463
29	570.6	18.2	861	11	BG420356
30	551.8	17.6	954	11	BF304259
31	551	17.6	560	10	AW411445
32	550.4	17.6	582	10	AI907882
33	529.8	16.9	627	10	AV729533
34	507	16.2	507	10	AI683520
35	506	16.1	757	10	BE282962
36	491.6	15.7	994	11	BT081073
37	489.6	15.6	592	11	BG078481
38	486	15.5	719	11	BG248068
39	478.4	15.3	939	11	BF316537
40	475.4	15.2	489	10	AL121548
41	463.2	14.8	612	10	AW129611
42	458.6	14.6	596	10	AI360995
43	456.4	14.6	1002	11	BG323892
44	452.6	14.4	918	11	BF178599
45	446.8	14.3	450	10	AI690075

ALIGNMENTS

RESULT 1

LOCUS BE746268 1122 bp mRNA EST 15-SEP-2000
DEFINITION 601579756F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3928572 5',
mrna sequence.
ACCESSION BE746268
VERSION BE746268.1 GI:10160260
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1122)
AUTHORS NIH-MGC <http://mgc.mci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM759 row: p column: 13
High quality sequence stop: 780.

FEATURES
source

Location/Qualifiers
1..1122
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3928572"
/clone.lib="NIH_MGC_9"
/tissue.type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:

Qy	679	atccggtttgcaccccgcccaagggacaggggaagcagcaagatcaatctggcggttgccaac	738
Db	242	ATCCGGTTGCACCCGCGCCCAAGACACAGGGAAGCAGAAATCAATCTGGCGGTTCGCAAC	301
Qy	739	acgggaattctagtgttttcagcggtttcactaaagatcaatgccttcaactgggccaaggtg	798
Db	302	ACGGGAATTCYTAGTGTTCAGGGTTTCACTAAGATCAATGCTTCACTGGGCCAAGGTG	361
Qy	799	cggaaagctgagcttcaagagggaagcgcttcttcataagactccggcgagatgcgaatagt	858
Db	362	CGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAGCTCCGGSCAGATGCCAATAGT	421
Qy	859	gcgtacacagataccttggaattctctgatgcccagtcgggattctgcaagtcctctctg	918
Db	422	GCGTACACAGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTCTGCAAGTCTTCTGG	481
Qy	919	aaaatcgtgtgaacatcatcgctctctttagactttttgaagagcccaaccaaagccc	978
Db	482	AAAATCTGTGTGAACATCATGCCCTCTTTAGACTTTTGAAGAGCCCAAAACCAAGCCC	541
Qy	979	aagccgcgtcctcttttagccgggggtcatcttgcgttcagtgctcgagactcagaagcag	1038
Db	542	AAGCCGCTCTCTTTAGCCGGGGTTCATCAITTCGGTTCACTGTCGGACTCAGAAGCAG	601
Qy	1039	gtctcagactatgtt--aaagaagagagacataagaagaagtgcaqgttgaagaagaagca	1096
Db	602	GTTCTCGACTATGTTTCACACAGGAAGGAGGACATAAGAAAGTGCAGTTTGAAGGAAGCACA	661
Qy	1097	gcaagattcatctatccgagccctgtctcacagctcacagaactgaatt-cggaagtg	1155
Db	562	G-AGATTCTATCTATCCGAGCCCTTGCTTACAGCCCTACAGAAGTGAATTCGGAAGTG	720
Qy	1156	ctggagcagctctcagcagagaccgaccttaccatttggagaggtgcgcaatctccaggg	1215
Db	721	CTGGAGCAGTCTCAGCAGAGCACCAGCTTACATTGTGAGAAGTGCCGAATCTCCAGGG	780
Qy	1216	ggccagagctcggcgagagaaag	1239
Db	781	GGACAGAGCTGCCGCGGAGGAAAG	804
RESULT	3		
BE260677			
LOCUS	BE260677	865 bp mRNA	EST 26-OCT-2000
DEFINITION	601154188F1 NIH_MGC_19	Homo sapiens cDNA clone	IMAGE:3510381 5'
ACCESSION	BE260677		
VERSION	BE260677.1	GI:9132166	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.G.E. Consortium/LLNL at: image.llnl.gov		
	Plate: LLCM194 row: 0 column: 22		
	High quality sequence stop: 730.		
FEATURES			
source	1..865		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3510381"		

/clone_lib="NIH_MGC_19"									
/tissue_type="neuroblastoma"									
/lab_host="DH10B (phage-resistant)"									
/notes="Organ: brain; Vector: pORF7; site_1: XhoI; site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."									
BASE COUNT	207 a	236 c	255 g	167 t					
ORIGIN									
Query Match	24.2%	Score 757.2;	DB 10;	Length 865;					
Best Local Similarity	97.3%;	Pred. No. 1.8e-175;							
Matches 823;	Conservative 0;	Mismatches 18;	Indels 5;	Gaps 5;					
Qy	671	tgtatggaatccggttgcaaccggcccaaggacaggggaagcacaagaatcaaatctgcccg	730						
Db	1	TGTATGGAATCCGTTGCAACCGGCCCAAGGACAGCAAGGCACGAATCAATCTGGCCG	60						
Qy	731	ttgccaacacaggaattctagtgtttcagggtttcactaagaatcaaatgccttaactggg	790						
Db	61	TTGCCAACACGGGAATTCTAGTGTTCAGGGTTTCACTAAGATCAATGCCCTTCAACTGGG	120						
Qy	791	ccaaggtcggaagctagcttcaaggaagcgtttctcatcaagctccgcgcagatg	850						
Db	121	CCAAGGTGCGGAAGCTGAGCTTCAAGGAAGACGCTTTCATCAAGCTCCGGCCAGATG	180						
Qy	851	ccaatagtgcgtaccaggaataccttggaaatcttcgtatggccagtcgggattctgcaggt	910						
Db	181	CCAATAGTGCGTACCAAGATACCTTTGGAATTCCTGTATGGCCAGTCGGGATTTCTGCAAGT	240						
Qy	911	cctctcgaaatctgtgtgaacatcatgcctctctttagacttttgaagagcccaaac	970						
Db	241	CTTTCTGGAAATCTGTGTGAACATCATGCTCTTTTAGACTTTTGAAGAGCCCAAC	300						
Qy	971	caaagcccaagccgctctctttagcgggggtcatcatcttcggttcagttgctcgactc	1030						
Db	301	CAAAGCCCAAGCCCGTCTCTTTAGCCGGGGTTCATCATTTTCGGTTTCAGTGGTGGACTC	360						
Qy	1031	agaagcaggttctcgactatgttaaagaagagagacataagaaggtgcagtttgaagga	1090						
Db	361	AGAAGCAGGTTCTCGACTATGTTAAAGAAGAGGACATAGAAGGTGCAGTTTGAAGGA	420						
Qy	1091	agcacagaagattcattctatccggagccttgcttcacagcctacagaactgaattcgg	1150						
Db	421	AGCACAGCAAGATTCTTCTATCGCGGAGCCCTGCTTTCACAGCCCTACAGAACTGAATTCGG	480						
Qy	1151	aagtgcctggagcagcttcagcagagcacacccttacatttgagaaaggtgccaaatctc	1210						
Db	481	AAGTCTGGAGAGCTCTCAGCAGAGCACACGCGCTTACATTTTGGAGAAGGTGCCGAATCTC	540						
Qy	1211	cagggggccagagctgccggcgaggaaggaacccgaaggtttccgcggggagccggggt	1270						
Db	541	CAGGGGCCAGAGTGC CGGC GAGAAAGGAACCAAGGTTTCCCGCGGAGCCGGGGT	600						
Qy	1271	cgcaaccgagccctgcgccgaggaagaccgcccggggttaaacaagcagcgagcagagccg	1330						
Db	601	CGCACCCGAGCCCTGCGCCGAGGAGAAGACCCCGCGGGTTAACAAAGCAGCGCAGGAGCGG	660						
Qy	1331	cctcggcgcccaacaggaagaagagagaggttcgittaaggaataggaccacagagagtaaac	1390						
Db	661	CCTCGCGCCCCACGGAGGAAGAGGAGGAGTCTG-TAAGGATAGGA-CCAGCAGAGATTAA	718						
Qy	1391	ctcagccccgcagcccaagcacaggtccctgact-ggcagtcctcaccttccagagctg	1449						
Db	719	CTCAGCCCCCCGAG-CAAGACACAGGTCCTCTGACTGGGCAGTCTCTCACTTTCGAGCTG	777						
Qy	1450	tctgtgaactcgaggggggagtgggccctcccaacgtgactgtctcccaactgagc	1509						

```
Db 778 TCTGTGACTCGCAGGGGAGTTGGCCCTCGGCCACGTAACC-TGTTTCCCAACCTGGGC 836
QY 1510 ccgac 1515
|| ||
Db 837 CCGAAC 842

RESULT 4
BG763918 784 bp mRNA EST 15-MAY-2001
LOCUS NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862444 5',
DEFINITION mRNA sequence.
ACCESSION BG763918
VERSION BG763918.1 GI:14074571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 784)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1722 row: o column: 21
High quality sequence stop: 761.
FEATURES
source
location
1..784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4862444"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DHL08 (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
ECORI; cDNA made by oligo-dT priming. Directionally cloned
into ECORI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library. |"
BASE COUNT 203 a 223 c 223 g 135 t
ORIGIN

Query Match 24.1%; Score 755.8; DB 11; Length 784;
Best Local Similarity 99.5%; Pred. No. 3.9e-175;
Matches 779; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1123 gcttcacagcctacagaactgaattcggaagtcgtggagcagtcctcagcagagcaccagc 1182
Db 2 GCTTCACGCCTACAGAACTGAATTCGGAAGTCTGGAGCAGTCTCAGCAGACACCAGC 61
QY 1183 cttacatttggaagagtgccgaattccaggggcccagagctgcgcgcgaggaaggaa 1242
Db 62 CTTACATTGGAGAGTGCCCGAATCTCCAGGGGGCCAGAGTCCCGCGCAGGAAGGAA 121
QY 1243 ccgaaggtttccgcggggagccgggtgcaccccgagccctgcgcgcgaggaagccccc 1302
Db 122 CCGAAGGTTTCCCGCGGGGAGCGGGGTGCACCCCGAGCCCTGCGCGCGAGGAAGCCCC 181
QY 1303 gcgggttaacagcagcgagcgagcgctcgccacagggaggaagaggaggagtc 1362
|||||
```

```
Db 182 GCGGGTAACAAGCAGCGCGAGCGCGCTCGGCCCCACGAGGAAGAGGAGGAGTC 241
QY 1363 gtaaggatagaccagcagagtaaacctcagcccccagcccccagcagcagcagcagcctccctg 1422
|||||
Db 242 GTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCCAGCCACGACAGGCTCCCTG 301
QY 1423 actggcagtcctcaccttcccgagctgtctgtgaactcaggggggagtgccccctgcc 1482
|||||
Db 302 ACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAACTCGACGGGGGAGTGGCCCCCTGCC 361
QY 1483 aacgtgacctgtctctccaaacctgagcccccgacacacaaagcagccctctcccttgatcagc 1542
|||||
Db 362 AACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCGCTCTCCCTTGATCAGC 421
QY 1543 ccgctgctgaatgaccagcgctgcgcccgagcgacgacgagcagatgagggcgaggaag 1602
|||||
Db 422 CCGCTGCTCAATGACGAGCGCTGCCCGGACGAGCATGAGGATGAGGCGCGGAGGAAG 481
QY 1603 agattcccaactgataaagcgtactctcatagctaagaagtgtctaccaccgagcagaaca 1662
|||||
Db 482 AGATTCCCAACTGATAAAGCGTACTTTCATAGCTAAGGAAGTGTCTACCACCGAGCGAACA 541
QY 1663 tatcgaagagatctcgaagtattcaacttgctggttttcagagcagcagtgacaaagggac 1722
|||||
Db 542 TATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCACAGTGAGCAAAAGAGGAC 601
QY 1723 gccatccggaagcactgaaagtctcatattcccgaaatttgaaactttgcacaaattt 1782
|||||
Db 602 GCCATGCCGGAAGCACTGAAAGTCTCATATTCGCCGAATTTGAACCTTTGCACAAATTT 661
QY 1783 catactaattttctcaagaaa-ttgagcaacgacttgcctgtggaagggcgctcaca 1841
|||||
Db 662 CATACTAATTTTCTCAAGGAAACTTGAGCAACGACTTGCCCTGTGGGAAGCGCGCTCAA 721
QY 1842 tgcccaaatcagagattaccagaatcgcgcatgcatgc-tgaagaacatttcagggca 1900
|||||
Db 722 TGCCCAATCAGAGATTACCAAGAAATCGCGCATGTCATGCTTGACGAACATTCAGGCCA 781
QY 1901 tga 1903
|||
Db 782 TGA 784

RESULT 5
AL566821/c 930 bp mRNA EST 16-FEB-2001
LOCUS AL566821 LTI_FLO13_Fbrnl Homo sapiens cDNA clone CS0DF025YJ01 3
DEFINITION prime, mRNA sequence.
ACCESSION AL566821
VERSION AL566821.1 GI:12919564
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
location/Qualifiers
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF025YJ01"
/dev_stage="pooled tissue from post conception fetuses (20
week_24 week and 26 week)"
/lab_host="DHL08"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
```


cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 186 a 249 c 283 g 211 t 1 others
ORIGIN

Query Match 24.0%; Score 753; DB 10; Length 930;
Best Local Similarity 97.9%; Pred. No. 2e-174;
Matches 806; Conservative 0; Mismatches 10; Indels 7; Gaps 4;

QY 2318 gqctccagcagcagtgcttcttcctgtcaacagcgtcctgtatcacagccggggc 2377
|||||
DB 930 GGCTCCAGCAGCGCATGTTCTTCCTTCAACAGCGTCTGTATACAGACCGGGGGC 871
|||||

QY 2378 tgacggctcccaatcagtttaagtcacagcgagctcccgctctatgcatgacgattg 2437
|||||
DB 870 TGACGGCTCCCAATCAGTTAAAGTCCAGGGCAGCTCCCGCTCTATGCGATGACGATTG 811
|||||

QY 2438 agagagcgaagacagagtg9gggggtgccccactgctgacctccgggcccagcgagtg 2497
|||||
DB 810 AGGAGAGCGAAGACAGAGT-GGGGGTGCCTCACTGCTGAACCTCCGGGGCCAGCGCGAGT 752
|||||

QY 2498 coactatcgtggcccgagctcgttcgagatgagagtgagtgagtgagtgagtgagtg 2557
|||||
DB 751 CCATCATGCTGGCCGCCAGTCTCGGTCCGAGATGAGAAAGTGGGTGAGGACATCCAGA 692
|||||

QY 2558 tggccattgacctgggagagagcagcagcccgccctgagttctctggccagcagcc 2617
|||||
DB 691 TGGCCATTGACCTGGGAGAGAGCAGCAGCCCCGCCCTGAGTCTCTGCGCAGCAGCC 632
|||||

QY 2618 cccctgacaaacagtccttgatgaagccacgcggtgaccagagtgatgagtgacc 2677
|||||
DB 631 CCCCTGACAAAGTCCCTGTATGAAGCACCACCGCGCTGACCAGGAGTCAGAGGATGACC 572
|||||

QY 2678 tgagcgtcgcgcacatcgtggagccagccagcccgccagccagccagccagccagccagcc 2737
|||||
DB 571 TGAGGCGCTTCGGCGACATCGCTGGAGCGCGAGG-CCCCGACCAGCGGACACAAATGGTGC 513
|||||

QY 2738 acgtgtgctggcaccgcaacacagcgtctccatggtgagttca---gcacgcagtg 2794
|||||
DB 512 ACGTGTGCTGGCACCACACACAGCGTCTCCATGGTGGACTTCACGCMCCGCCAGTGG 453
|||||

QY 2795 agaatacagttgtctggaacacctgctgaggaataatcaaaacagcaacggtgagcagaagc 2854
|||||
DB 452 AGAATCAGTTGTCTGGAACCTGCTGAGGAATTCAAAAACAGCAACGGGTGGCAGAAGC 393
|||||

QY 2855 tdtg--ggfgrgttcaaaacttgcgttcttcttcttcaaatcacaccagagacaatca 2912
|||||
DB 392 TGTGGCGTCGGTGTACAAAATCTTGCCTGTCTTCTTACAATCACACGAGCAATCA 333
|||||

QY 2913 tccctctgcagcctgcctctgctggactcgtcgtcaccatccccctctgagtcgagaa 2972
|||||
DB 332 TCCCTTGGCAGCCTGCCTCTGCTGGCTACTCGCTCACCATCCCTCTGTGTCGAGAA 273
|||||

QY 2973 catccagaagactacgtgttcaagctgcaactcaagtcceacgtctactacttcaggggc 3032
|||||
DB 272 CATCCAGAAGAACTACGTGTTCAAGTGCACATTCAAGTCCCAGTCTACTACTTCCAGGC 213
|||||

QY 3033 gaaaagcaggtacagttcgaaagtgatgaaatgattcagcagtgccacagctctgc 3092
|||||
DB 212 GGAACCGAGGTACAGTTCGAAGGTGGATGGAAGTGTATCCGAGTGCACCGAGTCTGC 153
|||||

QY 3093 ctccgacccccagctgttgagccacaagagctctcttggtat 3135
|||||
DB 152 CTCGGACCCACAGTGTGAGTTCACAAAGATCTCTTGTGTAT 110
|||||

RESULT 6

BG764061 861 bp mRNA EST 15-MAY-2001
LOCUS 602737071f1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862417 5',
DEFINITION mRNA sequence.
ACCESSION BG764061
VERSION BG764061.1 GI:14074714
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCFD/PTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1722 row: n column: 18
High quality sequence stop: 817.

FEATURES

source 1..861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4862417"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library. |"

BASE COUNT 198 a 245 c 270 g 147 t 1 others
ORIGIN

Query Match 23.8%; Score 745; DB 11; Length 861;
Best Local Similarity 96.9%; Pred. No. 1.8e-172;
Matches 791; Conservative 0; Mismatches 21; Indels 4; Gaps 3;

QY 1862 aaagaatcgccatgtcatgtgaagaacattcaggcagcagcagcagcagcagcagcagcagc 1921
|||||
DB 2 AAAGAATCGGCATGTCTATGCTGAGAAACATTCAGGGCATGAAGACCTTGGCGGCTCACCC 61
|||||

QY 1922 tgtgaagacacagcagcgcccttgaggccctggagaatggaatcaagagctcccggcgccg 1981
|||||
DB 62 TGTGGAAGCACAGCAGGCGCTTGGAGGCCCTTGAGAAATGAATCAAGAGCTCCCGCGCGC 121
|||||

QY 1982 tggagaactctgcagagagactttgagctgcagaagtggtttacctacccgctcaaacact 2041
|||||
DB 122 TGGAGAACTTCTGCAGAGACTTTTGTGCTGCAGAAAGGTGTGTACCTTACCGCTCAACACCT 181
|||||

QY 2042 tccctcgtcgccactgcacggctcatctacataacagagctctggagcagctgtgca 2101
|||||
DB 182 TCCTCTCGGGGCACCTGACCGGCTCATGCACCTACAAAGAGGTCTCTGGAGCGGCTGTGCA 241
|||||

QY 2102 aacacacccgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2161
|||||
DB 242 AACACCACCGCGGAGCCAGCGGACTTTCAGGGACTTCCGAGCGGCTNTGGCAGAGATCA 301
|||||

```
QY 2162 cggagatggtggcagagctccacggtacgatgatcgaatgaggaatttccagaagtgc 2221
Db 302 CGGAGATGTTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGC 361
QY 2222 acgaactcaagaagaatttgattggtgattgacaatttgggttcgggaaggagattca 2281
Db 362 ACGAATCTAAGAAGAATTTGATTGGCATTGACAATCTTGTGGTTCGGGAAGGAGTTCA 421
QY 2282 tccgttgggagcgtccacgaagctctcggggaagggtccagcagcgcatgttcttc 2341
Db 422 TCCGCTTGGCAGCCTCAGCAAGCTCTCGGGGAGGGGCTCCAGCAGCGCATGTTCTTC 481
QY 2342 tgttaacagactctgtctatacacagcggggggtgacggtcccaatcagtttaag 2401
Db 482 TGTTCACAGCTCTCTGTATACACGAGCGGGGCTGACGGCTCCAAATCAGTTTAAAG 541
QY 2402 tccacgggagctcccgctctatgcatgacgattgagagagcgaagcagagtgagg 2461
Db 542 TCCACGGGAGCTTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGAGTGGGG 601
QY 2462 tgcccacitgctgacccctccggggcagcggcagtcctcatcctggtggcagttctc 2521
Db 602 TGCCCCACTGCTGACCCCTCCGGGGCAGCGGCGAGTCCATCATCTGTGGCGCCAGTCTC 661
QY 2522 ggtccgagatggagaagtgggttgagacatccagatggccattgacctggcggaag- 2580
Db 662 GGTCCGAGATGGAGAAGTGGGTGAGGACATCCAGATGCCATTGACCTGGCGGAGAGA 721
QY 2581 agcagcagcccgccctgagttcttgccagcagcagcccccctgacaagaagtcctgat 2640
Db 722 AGCCACAGCCCCCGCTCAGTTCCTTGGCCAGCAGCCCCCTGACAAAGT-CCCTGAT 780
QY 2641 gaagcaccgcggtgaccagagtcagaggtgac 2676
Db 781 GAAGC--ACCGCGTGACAGGAGTCAGAGGATGAC 814

RESULT 7
BG475554
LOCUS 502491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5',
DEFINITION mRNA sequence.
ACCESSION BG475554
VERSION BG475554.1 GI:13407833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 901)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1381 row: h column: 11
High quality sequence stop: 818.
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4620346"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
```

```
BASE COUNT 232 a 243 c 253 g 173 t
ORIGIN
Query Match 23.2%; Score 726.2; DB 11; Length 901;
Best Local Similarity 91.3%; Pred. No. 7.8e-168;
Matches 817; Conservative 0; Mismatches 68; Indels 10; Gaps 4;
QY 1532 ccttgatcagccgcgtgctgaatgaccagcctgccccggagcagcagatgagtgag 1591
Db 2 CTTTGATGACGCCGTGCTGAATGACCAAGCTTGCCTCCCGGACGATGAGGATGAGG 61
QY 1592 gcccggaggaagattcccaactgataaagcgtacttcatagctaaaggaagtgtacc 1651
Db 62 GCGGAGGAAGAGATTCCCACTGATAAAGCGTACTTTCATAGCTAAGGAAGTGTCTACCA 121
QY 1652 ccgagcgaacatctatgaaaggtctcgaagtattatcacttcgtggttcagagcacagt 1711
Db 122 CCGAGCGAACAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCACAGTGA 181
QY 1712 gcaagagagacgcatccggaagcaactgaaaaagtctcatattcccgaaatttgaacct 1771
Db 182 GCAAGAGAGACCCATGCGGAGACTGAAAAGTCTCATATTTCCGGAATTTTGAACCTT 241
QY 1772 tgcaaaaatttatactaatcttcaaggaattgacaaagacttgcctgtgggaag 1831
Db 242 TGCACAAATTTCTACTAATTTCTCAAGAAATTCAGCAACGACTTGCCTGTGGGAAG 301
QY 1832 gccgtcaaatgccaaatcagagattaccaaagaatcgcgatgtcatctgcagaagaa 1891
Db 302 GCCGCTCAAAATGCCAAATCAGAGATTACCAAGAAATGCGCATGTCTGCTGAAGAACA 361
QY 1892 ttcaaggcatgaagcactgcccgtcactgtggaagcagcagcagccttgaggccc 1951
Db 362 TTCAGGGCATGAAGCAACCTTGGCGGCTCACCTGTGGGAAGCACAGCGGCTTGGAGGCC 421
QY 1952 tggagaatggaatcaagagctcccggcgctgagaaactctgcagagactttgagctgc 2011
Db 422 TGGAGAATGGAATCAAGAGCTCCCGCGCTGGAGAATCTCTGCAGAGACTTTGAGCTGC 481
QY 2012 agaagtggtgtacctaccgctcaacaccttctcctcgccgcaactgcaaccggtcatgc 2071
Db 482 AGAAGGTGTTTACCTTACCGCTCAACACCTTCTCTCGCGGCACCTGCACCGGCTCATGC 541
QY 2072 actcaagcagctcttgagcgtgtgcaaacaccaccccgagcagcagcagcttca 2131
Db 542 ACTAAGCAGGTCTCTGGAGCGGCTGTGCAACACCAACCGCGGACCGGACCGGCTTCA 601
QY 2132 -gggactcgcgagcgcgtttgagagatcacgagatggtggcacagctccacggtacg 2190
Db 602 GGGGACTGCGAGCGCGCTTTGGCAGAGATCACGGAGATGTTGGCAGAGCTCCACGGTACG 661
QY 2191 atgatacagatggagaatttccagaagctgcagaaactcaagaagaatttgatggcatt 2250
Db 662 ATGATCACGATGGAGAATCTCCAGAGCTGCCGAATCTCAAGACAAGATTTGATGGCAAT 721
QY 2251 gacaactctgtgtccgggaa- - - -ggaggttcatccctctgggcagcctcagca- - - 2303
Db 722 GACAATCTTGTGGTCCCGGGAAGCGGAGTTCCATCCGGTCTGCGGAGGCTTCAAGCAAGC 781
QY 2304 -gctctcgggggaagggtcccgagcagcgcatgttcttctctgttcaacagcgtcct-gcta 2361
Db 782 TCCTCGGGGAAAGGAGACTCCCGAGCGCATGTCTTCTTCTGTCACGAAGTCTCTGGCTT 841
QY 2362 tacacgagccgggggctgacggcctccaatcagtttaaaagtcacagcgagctcc 2416
|||||
```

ECORI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).*

```
Db 842 TACACAGCGGGCTGACAGGGCTCCATCCACGTTAAAGCCAAGGCAGTCCC 896

RESULT 8
LOCUS BF793662 736 bp mRNA EST 12-JAN-2001
DEFINITION 602254351f1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4346509 5',
mrna sequence.
ACCESSION BF793662
VERSION BF793662.1 GI:12098716
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9968 row: f column: 14
High quality sequence stop: 710.
FEATURES
    source
        1..736
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4346509"
            /clone_lib="NIH_MGC_84"
            /tissue_type="adrenal cortex carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:1:
            NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT
            primed. Average insert size 1.229 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT 167 a 228 c 209 g 132 t
ORIGIN

Query Match 22.8%; Score 716; DB 11; Length 736;
Best Local Similarity 99.2%; Pred. No. 2.4e-165;
Matches 730; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 2349 cgacgtctgtatcacgagcgggggctgacggcctccaatcagtttaaaagtcacgg 2408
Db 1 CGACGTCCTGCTATACACGAGCGGGGCTGACGGCTCCATCATCTTAAGTCCACGG 60
QY 2409 gcagctccgcgtctatggcatgacgattgagagagcgaagcagtgagggtgcacca 2468
Db 61 GCAGTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGAGTGGGGTGCCCCA 120
QY 2469 ctgcctgacctccgggagcggcagtcctcatcctggtggcgcagttctcggtccga 2528
Db 121 CTGCTGACCCCTCCGGGGCCAGCGGAGTCCATCATCTGTTGGCGCGCAGTTCGGTCCGA 180
QY 2529 gatgagaagtgggtgagacatccagatggccattgacctggcggagagaagcagcag 2588
Db 181 GATGGAGAGTGGTGTGAGACATCCAGATGGCCATTGACCTGGCGGAGAGAGCAGCAG 240
QY 2589 cccgcccctgagttctgtgccagcagccccctgacaaagtcacctgtatgaagccac 2648
Db 241 CCGCGCCCTGAGTTCCTGGCCAGCAGCCCCCTGTGACAAAGTCCCTCGATGATGAGCCAC 300
QY 2649 cgcggtgaccaggagtcagagatgacctgagcgcctcgccacatcgtctggagcgcca 2708
|||||
```

```
Db 301 CGGGGTACCAGGAGTACAGGATGACCTGACGGCTCGCGCATCATCTGCTGAGCGCCA 360
QY 2709 gggccgcaccgcggcacaacacaaatggtgcacgtgtgtgacccgcgaacacacagcgtctc 2768
|||||
Db 361 GGCCCGCACCGCGGCAACACAAATGTTGACAGTGTGCTGGCACCAGCAACACCGCTCTC 420
QY 2769 catgtggacttcagcatcgcaatggagaaatcaagtgtctgtgaaacctgtgaggaatt 2828
|||||
Db 421 CATGGTGGACTTTCAGCATCGAGTGGAGAAATCAGTTGTCTGGAAACCTGCTGAGGAAAT 480
QY 2829 caaaaacagacgcgggtggcagaagctgtgggtgtgttcacaaaacttctgcctgtctt 2888
|||||
Db 481 CAAAAACAGCAACGCGTGGCAGAGCTGTGGGTGGTTTCACAAACTTCTGCTGTTCTT 540
QY 2889 ctacaaatcacaccaggaacaatcatcccttgcacgctgcctctctgcgttactcgtc 2948
|||||
Db 541 CTACAAATCACACAGGACAATCATCCCTTGGCAGCTGCTGCTCGCTACTCGCT 600
QY 2949 caccatccctctgagtcgagaaacatccagaaagactaagtttcaagctgcaactcaa 3008
|||||
Db 601 CACCATCCCTCTGAGTCCGAGAACATCCAGAAAGACTAGCTGTTCAGGCTGCACTTCAA 660
QY 3009 gtccacgtctactacttcaggcggaagcaggtacacgcttcgaaggtggatgg-aag 3067
|||||
Db 661 GTCCACGCTCTACTACTTTCAGGGCGGAAGAGTACAGCTTCGAAGGTGGTGGCAAG 720
QY 3068 tgatccgcagtgccac 3083
|||||
Db 721 TGAATCCGAGTGCAC 736

RESULT 9
BE745887 898 bp mRNA EST 15-SEP-2000
LOCUS 601573513f1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3834699 5',
mrna sequence.
ACCESSION BE745887
VERSION BE745887.1 GI:10159879
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 898)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LLCM515 row: i column: 04
High quality sequence stop: 723.
FEATURES
    Location/Qualifiers
        1..898
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3834699"
            /clone_lib="NIH_MGC_9"
            /tissue_type="adenocarcinoma cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: ovary; Vector: pOTB7; Site:1: XhoI; Site:2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCAACAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using 2AP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT      236 a      229 c      269 g      164 t
ORIGIN
Query Match      22.5%; Score 705.4; DB 10; Length 898;
Best Local Similarity 96.8%; Pred. No. 1e-162;
Matches 762; Conservative 0; Mismatches 21; Indels 4; Gaps 4;
QY 613 ggcaaacaccagcagaatacagatttccagctcctagagattccgctcggttagagatg 672
DB 1 GGACAAACACCACGACGAGATATCCAGTCTCTAGAGATTCCGCTCGCTAGAGATG 60
QY 673 tatggaatccggttgccacccggccaaaggacagaggaagcagcaagatcaatctggccgtt 732
DB 61 TATGGAATCCGGTTGCACCCGGCCCAAGGACAGGAGGACGAAGATCAATCTGCGCGTT 120
QY 733 gccaaacacgggaattcttagtcttccaggtttcactaaagatcaatgccttcaactgggcc 792
DB 121 GCCAAACACGGGAATTCTAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCC 180
QY 793 aaggtcggaagctgagcttcaagagaaagcgtttctcatcaagctccgcccagatgcc 852
DB 181 AAGTCCGGAAGCTGAGCTTCAAGAGAAAGCGCTTCTCATCAAGCTCCGGCCAGATGCC 240
QY 853 aatagtcgtaccagataaccttggaattcctgatggcagtcgggatttcttcaagtcc 912
DB 241 AATAGTGCGTACCAAGATACCTTTGGAATTCCTGATGCCAGTCGGGATTTCTGCAAGTCC 300
QY 913 ttctggaatactgtgttgaacatcatgcttctttagactttttgaagagcccaacca 972
DB 301 TTTGTGAAAATCTGTG-TGAACATCATGCTCTTTTAGACTTTTTTGAAGAGGCCCAACCA 359
QY 973 aagcccaagccctcctcttagccgggggtatcatttcggttcagtggtcgactcag 1032
DB 360 AAGCCCAAGCCCGCTCTTTAGCCGGGGTCAATATTCGGTTCAGTGTCCGACTCAG 419
QY 1033 aagcaggttctcgactatgttaa-agaagaggacataagaaggtgcagtttgaaggaa 1091
DB 420 AAGCAGGTTCTCGACTATGTGTAACAGAAGGAGGACATAGAAGGTGCAGTTTGAAGGAA 479
QY 1092 gcacacaagattcattctatccgagccttgcttccacagcctacagaactgaattcgga 1151
DB 480 GCACAGCAAGATCTATCTATCCGAGCCTTGCTTCCACAGCCTACAGAACTGAATTCGGA 539
QY 1152 agtctggagcagctctcagcagagcaccagccttcatatttggagaaggtgcgaattctcc 1211
DB 540 AGTGCTGGACGACTCTCAGCAGAGCACCAGCCTTACATTTGGAGAGGTTGCCGAATCTCC 599
QY 1212 agggggccagagctgcgaggaaggaacccaagaggtttccgcccgggagccgggtc 1271
DB 600 A-GGGGCCAGAGCTGCGCGGAGGAAAGGAAACGAAAGGTTTCCGCGGGGAGCCGGGTC 658
QY 1272 gcaccagacccctgcgcagagagaagcccggttaacaagcagcagcagagagccgc 1331
DB 659 GCACCCAGCCCTTGCCCGAGGAGAAGCCCCCGGGGTAAACAGCAGCGGACGGAGCCG- 717
QY 1332 ctcgggccacaggaagagagaggttcgtttaaggaaggacccagcagagtaaac 1391
DB 718 CTCGGCCGCCACCGAGGAGAGAGGAGGCGCTTAAGATAGGACCCACAGAGAGTAAAC 777
QY 1392 tcagccc 1398
DB 778 TTAGCCC 784
RESULT 10
BG829192
LOCUS
DEFINITION
602753246F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4906114 5',
mRNA sequence.
ACCESSION
BG829192
VERSION
BG829192.1 GI:14176766
KEYWORDS
EST.
```

```
SOURCE          human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 835)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llni.gov
Plate: LLC1808 row: k column: 11
High quality sequence stop: 835.
FEATURES
Location/Qualifiers
1..835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4906114"
/tissue_type="NIH_MGC_17"
/lab_host="rhabdomyosarcoma"
/notes="Organ: muscle; Vector: pORF7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life technologies)."
```

BASE COUNT 213 a 231 c 243 g 148 t
ORIGIN
Query Match 22.5%; Score 704.6; DB 11; Length 835;
Best Local Similarity 95.8%; Pred. No. 1.6e-162;
Matches 800; Conservative 0; Mismatches 24; Indels 11; Gaps 7;
QY 928 gttgaacatcatgctcttcttagacttttgaagagcccaaacgaagccagccgcgc 987
DB 2 GTTGAACATCATGCTTCTTTA-ACTTTGTGAAGAGCCCAACCAAGCCCAAGCCGTC 60
QY 988 ctctttagccgggggtcatttcggttcagtcggttcgactcagaagcaggttctcgac 1047
DB 61 CTCTTTAGCCGGGGTGCATCATTTCCGTTTCAGTGTGCGACTCAGAAGCAGGTCTCGAC 120
QY 1048 tatgttaagaagagagacataagaaggtgcagtttgaaggaagcacagcaagattcat 1107
DB 121 TATGTTAAAGAGGAGGACATAGAAGGTGCAGTTTGAAGAGGAGGACACCAAGATTCAT 180
QY 1108 tctatccgagccttcttcacagcctacagaactgaattcggaagtgcgtggagcagctc 1167
DB 181 TCTATCCGAGCCTTCTTTCACAGCCTACAGAACTGAATTCGGAAGTGTGTGGAGCAGTCT 240
QY 1168 cagcagagcaccagccttaccatttgagaaggtgcgcgaatctccagggggccagagctgc 1227
DB 241 CAGCAGAGCACCAGCCTTACATTTGGAGAAGGTGCCGAATCTCCAGGGGGCCAGAGCTGC 300
QY 1228 cggcgaggaagaaccgaaggtttccgcccgggagccgggttcgcaccgcagccctgcg 1287
DB 301 CGCGGAGGAAGAACCCGAGAGGTTTCCGCGGGAGGCCCGGGGTCCGACCCAGCCCTCGC 360
QY 1288 ccgaggaagcccccggggtaacaagcagggcgagcagcagccctcgcgcgccacaggag 1347
DB 361 CCGAGGAGAAGCCCCGCGGGTAAACAAGCAGGCGGACGAGCGCCTCGGCGCCACCGGAG 420
QY 1348 gaagagagagaggtcgtt-aaggataggagccagagagtaaacctcagcccccgagcc 1406
|||||

```
Db 421 GAAGAGGAGGAGTCTTCAAGGATAGGACCCAGACAGAGTAAACCTCAGCCCCCGCAGCC 480
Qy 1407 aagcacagctccctgactgagctcctcaacttccagctgtctgtgaactcgcagg 1466
Db 481 AAGCAGAGGCTCCCTGACTGGCAGTCTCCTACCTTTCCGAGCTGTCTGTGAATCGCAGG 540
Qy 1467 gggagtgccctgccaagtgacctgtctccaaactgagcccgacacacaaagcaggc 1526
Db 541 GGGAGTGGCCCTGCCAAGTACCTTGTCTCCAGACTGAGCCCCGGAACCAAGCAGGCG 600
Qy 1527 ctctcccttgatcagccgcgtgc-tgaatgaccagcctgccccggagcgagcagtgagg 1585
Db 601 CTCCTCCCTTGATCAGCCCGCTGTTGAATGACCAAGCCTGCCCGCGGAGCAGATGAGG 660
Qy 1586 atgagggcgaggaagagattcccaactgataagcgtacttca-tagctaagaagt 1644
Db 661 ATGAGGGCGGAGGAAGAGATTCCCAACTGATAAAGCGTACTTCAATTAGCTAAGGAAGTG 720
Qy 1645 tctaccacgagcgaac--atatctgaagatctcgaag-ttatcactctgttttca 1700
Db 721 TCTACCAAGGAGCAACCATATCTGAGGGGATCTCGAAGTTTATCACTTCCGGGTTCAG 780
Qy 1701 gagcacagtgaagcagagggagccatcccggaagcac---tgaagaagtctcata 1752
Db 781 AAGCACAGTGAAGCAAGGAGGACGCCATGCCCGGAAGACACTTGAAGAAGTCTCATA 835

RESULT 11
BE910036 601498085F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3900021 5',
LOCUS mRNA sequence.
DEFINITION BE910036 902 bp mRNA EST 20-OCT-2000
VERSION BE910036.1 GI:10406227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9698 row: j column: 22
High quality sequence stop: 750.
FEATURES
source
Location/Qualifiers
1..902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3900021"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site1: NotI;
Site2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 231 a 248 c 272 g 151 t
ORIGIN

Query Match 22.3%; Score 700.2; DB 11; Length 902;
Best Local Similarity 95.8%; Pred. No. 1.9e-161;
Matches 752; Conservative 0; Mismatches 28; Indels 5; Gaps 3;
```

```
Qy 1143 gaattcggaagtgcctggagcagctcagcagagcaccagccttaccatttggagaaggtgc 1202
Db 1 GAATTTCGGAAGTGCCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTCGGAAGGTGC 60
Qy 1203 cgaatctccagggcgccagagctgccggcgaggaaggaaccgaaggtttccgcgcggga 1262
Db 61 CGAATCTCCA-GGGCCAGAGCTGCCGGCGAGGAAGGAACCGAAGTTCGCCCGGGGA 119
Qy 1263 gccggggtgcaccgcagcctgcgcggaggaagccccgcgggtaacaaagcagcgga 1322
Db 120 GCCGGGTGCGACCCGAGCCCTGCGCCGAGGAGAACCCCGGGGTAAACAAGCAGCGGA 179
Qy 1323 cggagccgctcggcgccacggaggaagagagaggtcggttaagatagagccagca 1382
Db 180 CCGAGCCGCTCGCGGCCACCGAGGAGGAGGAGGTTCGTTAAGATAGGAGCCAGCA 239
Qy 1383 ggtaaacctcagccccgcagcgaagcagcagcctcctgactggcagtcctcacccttc 1442
Db 240 GAGTAAACCTCAGCCCCCGCAGCAAGCACAGGCTCCCTGACTGGCAGTCTCACCTTTC 299
Qy 1443 cgagctgtctgtgaactcgagggggagtgcccctgcgaactgacettgtctcccaa 1502
Db 300 CGAGCTGTCTGTGAATCGCAGGGGGAGTGGCCCTGCGCAACGTGACCTTGTCTCCCAA 359
Qy 1503 cctgagccccgacacaaagcagcctcctccttgatcagccccgcgtcgtgaatgaccagc 1562
Db 360 CCTGAGCCCCGACACCAAGCAGGCCCTCTCCCTTGATCAGCCCGCTGCTGAATGACCAGGC 419
Qy 1563 ctgccccgcgagcagcagtgagtgagggcgaggaagagagattcccaactgataaagc 1622
Db 420 CTGCCCCCGGAGCGGACGATGAGATGAGGGCCGCGAGGAAGAGATTCCCAACTGATTAAGC 479
Qy 1623 gtactctatgctaaggaagtgtctaccaccgagcgaacatatctctgaagatctcgaagt 1682
Db 480 GTACTTCATAGTAAAGGAAGTGTCTACCAACCGAGCGCAACATATCTCAAGGATCTCGAAGT 539
Qy 1683 tatcactctggttttcagagcacagtgagcaaaaggagcgcctatcccg-aaagcactga 1741
Db 540 TATCATCTTCGTGGTTTCAGAGCACAGTGAAGAAAGAGAGCGCCATCCCGGAAGACATGA 599
Qy 1742 aaa---gtctcatattccgaatttgaacctttgcacaaattctcactaatcttctca 1798
Db 600 AAAAGTCTCATATTTCCTCGAATTTTGAAACCTTTGACAAAATTTCTACTAATTTCTCA 659
Qy 1799 aggaattgacacagcactgcccgtgtgggaaggcgctcaaatgcccaaatcagagatt 1858
Db 660 AGGAATTTGAGCAACGACTTGCCCTGTGGGAAGGCGGTCAAAATGCCCAATCAGAGATT 719
Qy 1859 accaaagaatcgcgatgtcatgtgtaagaacattcagggcgatgaagcaccctggcgctc 1918
Db 720 ACCAAAGAATCGCGATGTCTGCTGAAGAACCATTTCAGCCCATAAAGCACCAGCGGGGTCC 779
Qy 1919 acctg 1923
Db 780 CTGTG 784

RESULT 12
AUI32546
LOCUS AUI32546 NT2RP4 Homo sapiens cDNA clone NT2RP4000050 5', mRNA
DEFINITION sequence.
ACCESSION AUI32546
VERSION AUI32546.1 GI:10992900
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 721)
Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y., Saito, K., Yamamoto, J.,
Nishikawa, T., Nakamura, Y., Nagai, T., Sugano, S., Masuho, Y. and
```



```
Db 717 GTGCTGAGCAGTCTCAGCAGAGCACCAGGCTTACATTGGAGAGGTGCCGAATCTCCA 776
QY 1213 gggggccagctgcccgcagagaaagaccgaagttttccgcgggagccgggtcg 1272
Db 777 -GGGGCCAGAGCTG-CGGCAGAGAAAGAACGAGGTTCCGGCGGAGCGGGTCGACCCG 834
QY 1273 caccgagcc 1282
Db 835 AGCCTGCGCC 844

RESULT 15
LOCUS BF206296
DEFINITION BF206296.1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4100044 5',
            mRNA EST 06-NOV-2000
ACCESSION BF206296
VERSION BF206296.1 GI:11099882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 883)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: c9apbs@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM969 row: i column: 05
            High quality sequence stop: 722.

FEATURES
            Location/Qualifiers
                1..883
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4100044"
                /clone_lib="NIH_MGC_19"
                /tissue_type="neuroblastoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCACGAG(G). Library constructed by Ling Hong
                in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
BASE COUNT 192 a 264 c 266 g 161 t
ORIGIN
```

Query Match 21.5%; Score 673; DB 11; Length 883;
Best Local Similarity 94.7%; Pred. No. 9.5e-155;
Matches 772; Conservative 0; Mismatches 35; Indels 8; Gaps 7;

```
QY 2298 cagcaagctctcgggaaagggtccagcagcagctgttctcttcaacagacgtcct 2357
Db 2 CAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTCCGTTCACGACGTCT 61
QY 2358 gctatacagcgggggggtgacggcctccaatcagtttaagtccacggcagctccc 2417
Db 62 GCTATACAGAGCGGGGGGCTGACGGCTCCCAATCAGTTAAAGTCCACGGGCGAGTCCC 121
QY 2418 gctctatgcatgacgattgagagagcagcagcagtgagggtgggggtgcccactgctgac 2477
Db 122 GCTCTATGCGATGACGA-TGAGAGAGCGGAAGACGAGTGGGGGGTCCCCCACTGCCTGAC 180
```

Search completed: December 6, 2001, 10:35:03
Job time: 7035 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	3442	100.0	3442	20	AAx79183	Human chondrocyte-
2	3437.2	99.9	4687	21	AAA08582	Human cytoskeleton
3	1688.6	49.1	3094	21	AAC98992	Human pancreatic c
4	223.6	6.5	241	21	AAA42150	Human secreted exp
5	169.6	4.9	3620	21	AAA08581	Human cytoskeleton
6	151.6	4.4	544	22	AAH98023	Murine 7-transmemb
7	144	4.2	3984	18	AAT58627	Protein tyrosine p
8	144	4.2	3984	20	AAx78463	Human PTPH1 CDNA.
9	143.2	4.2	2872	22	AAS08887	Human protein tyro
10	133.6	3.9	3166	22	AAH17787	Human cDNA sequenc
11	95.8	2.8	659	22	AAH33845	Human colon cancer

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

SUMMARIES

PS Claim 5; Fig 1; 59pp; Japanese.

XX This sequence represents the coding region for a protein (CDEP) expressed
CC in differentiated human foetal chondrocytes, which contains an exrin-like
CC domain, a DBI homology (DH) domain and a pleckstrin homology (PH) domain.
CC The nucleic acid or protein can be used in the investigation and
CC treatment of cancers and arthritic diseases (including chronic rheumatoid
CC arthritis), or for screening of candidate anticancer drugs.

XX Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;

Query Match		100.0%;	Score 3442;	DB 20;	Length 3442;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 3442;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	cgccgagccgcccgcgtgtggagatattctctaaagccgtttcatcatcattgaggagaata	60		
DB	1	cgccgagccgcccgcgtgtggagatattctctaaagccgtttcatcatcattgaggagaata	60		
QY	61	gagcagagccgacccagagatcacgactggggcccgcaaaaattcgggatacgtacc	120		
DB	61	gagcagagccgacccagagatcacgactggggcccgcaaaaattcgggatacgtacc	120		
QY	121	ttgaaactgtgacagaagccgcccccaacaccttcagaaactcgtgtccatcaaatc	180		
DB	121	ttgaaactgtgacagaagccgcccccaacaccttcagaaactcgtgtccatcaaatc	180		
QY	181	cagatgtgtagtacacccagaggaatttgaagtctccaaagagctcctgggaagtgtg	240		
DB	181	cagatgtgtagtacacccagaggaatttgaagtctccaaagagctcctgggaagtgtg	240		
QY	241	ctgtgtgtagcagtttgcacacacctcaactcgtggaagtgtgactattttggcctcgtg	300		
DB	241	ctgtgtgtagcagtttgcacacacctcaactcgtggaagtgtgactattttggcctcgtg	300		
QY	301	tttctgtatcacaaaaagatacacggtgtggtgtggtatctctaaacaccttggaaacag	360		
DB	301	tttctgtatcacaaaaagatacacggtgtggtgtggtatctctaaacaccttggaaacag	360		
QY	361	attagaagccaaagacggtgtgttaagtgtgtggtgaaattcttcgcctgaccac	420		
DB	361	attagaagccaaagacggtgtgttaagtgtgtggtgaaattcttcgcctgaccac	420		
QY	421	acaaactccaaagaagaaactcaaggtacactgtctgcgtgcaggtggaagcagacttg	480		
DB	421	acaaactccaaagaagaaactcaaggtacactgtctgcgtgcaggtggaagcagacttg	480		
QY	481	gctcaagggaggttgacgtgtaatagacacagcgcagctctctgtatttcacacattgtg	540		
DB	481	gctcaagggaggttgacgtgtaatagacacagcgcagctctctgtatttcacacattgtg	540		
QY	541	caactcagactgggatttggatgaagccttgagacagagcacttagcaaaaaataaa	600		
DB	541	caactcagactgggatttggatgaagccttgagacagagcacttagcaaaaaataaa	600		
QY	601	tacatactcagcaagcactagagcaaaaaactgtggaatttcaccataaacacatt	660		
DB	601	tacatactcagcaagcactagagcaaaaaactgtggaatttcaccataaacacatt	660		
QY	661	ggacaaacacacagagaatacagatttccagctctcagatagattggccctcgttagagatg	720		
DB	661	ggacaaacacacagagaatacagatttccagctctcagatagattggccctcgttagagatg	720		
QY	721	tatggaatccggttgcacccggccaaaggacaggaaggaagcagaagatcaatctggccgtt	780		
DB	721	tatggaatccggttgcacccggccaaaggacaggaaggaagcagaagatcaatctggccgtt	780		
QY	781	gccaaacacgggaattcttagttttcaggttttcaactaagatcaatgcttcaactgggcc	840		
DB	781	gccaaacacgggaattcttagttttcaggttttcaactaagatcaatgcttcaactgggcc	840		
QY	841	aaggtgcggaagctgagcttcaagagagaagcgttttctcatcaagctccggccagatgccc	900		

DB	841	aaggtgcggaagctgagcttcaagagagaagcgtttctcatcaagctccggccagatgcc	900		
QY	901	aatagtcgtaccaggaataccttgaattctctgatgcccagtcgggatttctgcaagtcc	960		
DB	901	aatagtcgtaccaggaataccttgaattctctgatgcccagtcgggatttctgcaagtcc	960		
QY	961	ttctggaaaatctgtgtgaaacatcatgctctcttttagactttttgaagagcccaacca	1020		
DB	961	ttctggaaaatctgtgtgaaacatcatgctctcttttagactttttgaagagcccaacca	1020		
QY	1021	aagcccaagccgctctctttagccgggggtcatcattcgggttcagtggtcggaactcag	1080		
DB	1021	aagcccaagccgctctctttagccgggggtcatcattcgggttcagtggtcggaactcag	1080		
QY	1081	aagcaggtttctgactactgttaaagaagaggagacataagaaggtgcaagttgaaaggaag	1140		
DB	1081	aagcaggtttctgactactgttaaagaagaggagacataagaaggtgcaagttgaaaggaag	1140		
QY	1141	cacagaagaattcattctatccggagccttgcctcacagcctcacagaactgaattcggaa	1200		
DB	1141	cacagaagaattcattctatccggagccttgcctcacagcctcacagaactgaattcggaa	1200		
QY	1201	gtgtgagcagttctcagcagagacacacagccttacatttggaaggtgcccgaattctcca	1260		
DB	1201	gtgtgagcagttctcagcagagacacacagccttacatttggaaggtgcccgaattctcca	1260		
QY	1261	gggggccaagctgcgcgcgaggaagaaacacgaaggtttccgcggggagccggggtcg	1320		
DB	1261	gggggccaagctgcgcgcgaggaagaaacacgaaggtttccgcggggagccggggtcg	1320		
QY	1321	caaccgagcctgcgcgcgaggaagacccgcgcgcgttaacaagcagcgagcgagcgcc	1380		
DB	1321	caaccgagcctgcgcgcgaggaagacccgcgcgcgttaacaagcagcgagcgagcgcc	1380		
QY	1381	tcgggccccgcgaggaagaggaggttcgttaaggataggaccccgagagtaaacct	1440		
DB	1381	tcgggccccgcgaggaagaggaggttcgttaaggataggaccccgagagtaaacct	1440		
QY	1441	cagccccgcagcaaacacacagcctcctgactgagcagtcctacatttccgagctctct	1500		
DB	1441	cagccccgcagcaaacacacagcctcctgactgagcagtcctacatttccgagctctct	1500		
QY	1501	gtgaactcgcagggggagtgccctgcgaacgtgacctgtctcccaacctgagcccc	1560		
DB	1501	gtgaactcgcagggggagtgccctgcgaacgtgacctgtctcccaacctgagcccc	1560		
QY	1561	gacaccaagcagggcctctccttgatcagcccgcgtgctgaatgacacagggcctgccccgg	1620		
DB	1561	gacaccaagcagggcctctccttgatcagcccgcgtgctgaatgacacagggcctgccccgg	1620		
QY	1621	acggacgatgagatgagggccgaggaagagattcccaactgataaagcgtacttcata	1680		
DB	1621	acggacgatgagatgagggccgaggaagagattcccaactgataaagcgtacttcata	1680		
QY	1681	gctaagaaagtgtctaccacccgagcaacatactgaaggtatctcgaagtattacactcg	1740		
DB	1681	gctaagaaagtgtctaccacccgagcaacatactgaaggtatctcgaagtattacactcg	1740		
QY	1741	tggtttcagagcagctgagcaaaagagacgccaatgcccgaagcactgaaaaagtcctata	1800		
DB	1741	tggtttcagagcagctgagcaaaagagacgccaatgcccgaagcactgaaaaagtcctata	1800		
QY	1801	ttccccgaatttgaacaccttgcaaaatttctactactaattttctcaaggaaattgagcaa	1860		
DB	1801	ttccccgaatttgaacaccttgcaaaatttctactactaattttctcaaggaaattgagcaa	1860		
QY	1861	cgacttgcctgtgggaagccgctcaaatgcccnaatcagagattaccnaaagaatcgcc	1920		
DB	1861	cgacttgcctgtgggaagccgctcaaatgcccnaatcagagattaccnaaagaatcgcc	1920		
QY	1921	gagtgtatgctgagaacacatttcaggggcatgaagcactgcccgtcactctgtggaagcac	1980		

Db 1921 gatgtcatgctgaagaacattcaggcgcatgaagcacctgaggctcacctgtggaagcac 1980
QY 1981 agcagagccttgagccctgagaatgaatacgaagctcccgcggtggaagacttc 2040
Db 1981 agcagagccttgagccctgagaatgaatacgaagctcccgcggtggaagacttc 2040
QY 2041 tgcagagacttgagctgcagaaggtgtgttacctaccctcaaaccttctctcgcg 2100
Db 2041 tgcagagacttgagctgcagaaggtgtgttacctaccctcaaaccttctctcgcg 2100
QY 2101 ccaactgcacggctcatgactacaagcaggttctctgagcggtgtgcaaacacaccccg 2160
Db 2101 ccaactgcacggctcatgactacaagcaggttctctgagcggtgtgcaaacacaccccg 2160
QY 2161 ccgagccacgcgaacttcaggactcccgagccgcttggcagagatcacggagatgggtg 2220
Db 2161 ccgagccacgcgaacttcaggactcccgagccgcttggcagagatcacggagatgggtg 2220
QY 2221 gcacagctccacgggtacgatgatcaagatggagaatttcagaagctgcacgaactcaag 2280
Db 2221 gcacagctccacgggtacgatgatcaagatggagaatttcagaagctgcacgaactcaag 2280
QY 2281 aaagatttgattggcattgacaatcttgggtccgggaaggaggttcacgtctgggac 2340
Db 2281 aaagatttgattggcattgacaatcttgggtccgggaaggaggttcacgtctgggac 2340
QY 2341 agcctcagcaagctctcggggaagggtcccgagcgcgatgttcttctgttcaacgac 2400
Db 2341 agcctcagcaagctctcggggaagggtcccgagcgcgatgttcttctgttcaacgac 2400
QY 2401 gtctctatacacagcagcggggtgacggcctcaaatcagtttaaaagtccaacggcgag 2460
Db 2401 gtctctatacacagcagcggggtgacggcctcaaatcagtttaaaagtccaacggcgag 2460
QY 2461 ctcccgctctatggcattgacgattgagagcagcagagcagagtggtggtcccaactgc 2520
Db 2461 ctcccgctctatggcattgacgattgagagcagcagagcagagtggtggtcccaactgc 2520
QY 2521 ctgacctccggggccagcggcagtcacatcatctggtgcgcagctctcgttcgagatg 2580
Db 2521 ctgacctccggggccagcggcagtcacatcatctggtgcgcagctctcgttcgagatg 2580
QY 2581 gagaagtgggtgagagatccaagatggccattgacgtggcgagagaagcagcagcccc 2640
Db 2581 gagaagtgggtgagagatccaagatggccattgacgtggcgagagaagcagcagcccc 2640
QY 2641 gcccttgattctctgccagcagccccctgacaacaagtccctgatgaagccaccccg 2700
Db 2641 gcccttgattctctgccagcagccccctgacaacaagtccctgatgaagccaccccg 2700
QY 2701 gctgaccagagtcagagatgacctgagcgcctcgcacatcgtgtgagcgcagggcc 2760
Db 2701 gctgaccagagtcagagatgacctgagcgcctcgcacatcgtgtgagcgcagggcc 2760
QY 2761 ccgcacccgggcaacaataatgtgacgtgtgctggtgcacgcgaacacccagcgtctccatg 2820
Db 2761 ccgcacccgggcaacaataatgtgacgtgtgctggtgcacgcgaacacccagcgtctccatg 2820
QY 2821 gtggacttcagatcgcagtgagagatcagttgtctggaacccctgctgaggaattcaaa 2880
Db 2821 gtggacttcagatcgcagtgagagatcagttgtctggaacccctgctgaggaattcaaa 2880
QY 2881 aacagcaacgggtggcagaagctgtgggtgtgttcacaaacttgcctgttcttctac 2940
Db 2881 aacagcaacgggtggcagaagctgtgggtgtgttcacaaacttgcctgttcttctac 2940
QY 2941 aaatcacaccaggacaataatccccctggcagcctcctctgctcgtcgtactgcctcacc 3000
Db 2941 aaatcacaccaggacaataatccccctggcagcctcctctgctcgtcgtactgcctcacc 3000
QY 3001 atccccctctgagtcggagaaatccagaaagactacgtgttcaagctgcacttcaagtcc 3060
Db 3001 atccccctctgagtcggagaaatccagaaagactacgtgttcaagctgcacttcaagtcc 3060

QY 3061 cacgtctactacttcaggcggaagcggagtcacacgttcgaaaaggtggatggaagtgtac 3120
Db 3061 cacgtctactacttcaggcggaagcggagtcacacgttcgaaaaggtggatggaagtgtac 3120
QY 3121 ccagtgccacacagctctcctcgcgcacccacggtgttgagccacaaagagtcctctgtg 3180
Db 3121 ccagtgccacacagctctcctcgcgcacccacggtgttgagccacaaagagtcctctgtg 3180
QY 3181 tattgatggccggacacactcgtttccgcagtggtgcttctcctgaaagacgtttccttt 3240
Db 3181 tattgatggccggacacactcgtttccgcagtggtgcttctcctgaaagacgtttccttt 3240
QY 3241 cttctgtattaatgaagcctggttaaaattaaacacctgtctgaaaaatacaaaacatggctt 3300
Db 3241 cttctgtattaatgaagcctggttaaaattaaacacctgtctgaaaaatacaaaacatggctt 3300
QY 3301 ccacagcagctctcctgtctccacagcgcggtttttaaccccgacctctcagcgtttgaa 3360
Db 3301 ccacagcagctctcctgtctccacagcgcggtttttaaccccgacctctcagcgtttgaa 3360
QY 3361 tgaacagcgtcccaacctccagtcctcctggcatccgctggggcgctgttctttagctagt 3420
Db 3361 tgaacagcgtcccaacctccagtcctcctggcatccgctggggcgctgttctttagctagt 3420
QY 3421 ccagttataaaacattgtcatt 3442
Db 3421 ccagttataaaacattgtcatt 3442

RESULT 2

AAA08582

ID AAA08582 standard; DNA; 4687 BP.

XX

AC AAA08582;

XX 19-JUL-2000 (first entry)

XX

DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.

XX

KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;

KW autoimmunity; inflammatory; vesicle trafficking; neurological;

KW cardiovascular; cell motility; reproductive; muscle disorder; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 90..3227

FT /*tag= a

FT /product= CYSKP-2

XX

PN WO200017355-A2.

XX

PD 30-MAR-2000.

XX

PF 17-SEP-1999; 99WO-US21565.

XX

PR 18-SEP-1998; 98US-0172226.

XX

PR 27-APR-1999; 99US-0131321.

XX

PA (INCY-) INCYTE PHARM INC.

XX

PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;

PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;

XX

DR WPI: 2000-283582/24.

DR P-PSDB; AAY91947.

XX

XX Human cytoskeleton associated proteins, used to treat cell

XX proliferative, autoimmune/inflammatory, vesicle trafficking,

XX neurological, cell motility, reproductive and muscle disorders

PS Claim 9; Page 101-102; 113pp; English.

QY 1861 cgacttgccttgtagaaggccgctcaaatgcccaaatcagaagattacaaagaatcggc 1920
|||||
Db 1902 cgacttgccttgtagaaggccgctcaaatgcccaaatcagaagattacaaagaatcggc 1961
QY 1921 gatgtcatgctgaagaacatcagggcatgaagcacctggcggctcacctgtggaagcac 1980
|||||
Db 1962 gatgtcatgctgaagaacatcagggcatgaagcacctggcggctcacctgtggaagcac 2021
QY 1981 agcagggccttgagggccctggagaatggaatcaagagctcccggcgctggagaacttc 2040
|||||
Db 2022 agcagggccttgagggccctggagaatggaatcaagagctcccggcgctggagaacttc 2081
QY 2041 tgcagagactttgagctgcagaaggctgtgttacctacacgctcaaacacttctcctcggg 2100
|||||
Db 2082 tgcagagactttgagctgcagaaggctgtgttacctacacgctcaaacacttctcctcggg 2141
QY 2101 ccaetgacccggctcatgcactacaagcagctcctggagcggctgtgcaaacaccaccgg 2160
|||||
Db 2142 ccaetgacccggctcatgcactacaagcagctcctggagcggctgtgcaaacaccaccgg 2201
QY 2161 ccgagccacgcgacttcagggactgcgagccgctttggcagagatcacggagatggtg 2220
|||||
Db 2202 ccgagccacgcgacttcagggactgcgagccgctttggcagagatcacggagatggtg 2261
QY 2221 gcaagctccacggctacgatgataaagatgagaatttccagaagctgcagaaactcaag 2280
|||||
Db 2262 gcaagctccacggctacgatgataaagatgagaatttccagaagctgcagaaactcaag 2321
QY 2281 aaagatttgattggcatgacaactctgtgttcgggaaaggagttcatcgtctgggc 2340
|||||
Db 2322 aaagatttgattggcatgacaactctgtgttcgggaaaggagttcatcgtctgggc 2381
QY 2341 agctcagcaagctctcgggaaagggtccagcagcagcatgttcttctgttcaacgac 2400
|||||
Db 2382 agctcagcaagctctcgggaaagggtccagcagcagcatgttcttctgttcaacgac 2441
QY 2401 gtcctgtatcacagcagccggggtcgcgcgcctccaatcagtttaaaagtcacaggcag 2460
|||||
Db 2442 gtcctgtatcacagcagccggggtcgcgcgcctccaatcagtttaaaagtcacaggcag 2501
QY 2461 ctccgcctctatggcatgacatgagagatgagagcagagcagtgagggggtgcccactgc 2520
|||||
Db 2502 ctccgcctctatggcatgacatgagagatgagagcagagcagtgagggggtgcccactgc 2561
QY 2521 ctgacctccggggccagcgcagtcctcatcgttgcgcgcagttctcgttccgagatg 2580
|||||
Db 2562 ctgacctccggggccagcgcagtcctcatcgttgcgcgcagttctcgttccgagatg 2621
QY 2581 gagaagtgggttaggacatccagatggccattgacctggcgggagaagagcagcagcccc 2640
|||||
Db 2622 gagaagtgggttaggacatccagatggccattgacctggcgggagaagagcagcagcccc 2681
QY 2641 gcccttgatttcttgccagcagccccctgacaaagtcctctgatgaagccaccgg 2700
|||||
Db 2682 gcccttgatttcttgccagcagccccctgacaaagtcctctgatgaagccaccgg 2741
QY 2701 gctgaccaggagtcagagatgacctgagcgcctcgcgcacatcgtgagcgcagggcc 2760
|||||
Db 2742 gctgaccaggagtcagagatgacctgagcgcctcgcgcacatcgtgagcgcagggcc 2801
QY 2761 ccgacccgggcaacacaaatggtgcacgtgtgctggcaccgcaaacaccagcgtctccatg 2820
|||||
Db 2802 ccgacccgggcaacacaaatggtgcacgtgtgctggcaccgcaaacaccagcgtctccatg 2861
QY 2821 gtggacttcagcatcgcagtgagaaatcagttgtctggaacctgtgagaaattcaaa 2880
|||||
Db 2862 gtggacttcagcatcgcagtgagaaatcagttgtctggaacctgtgagaaattcaaa 2921
QY 2881 aacagcaacgggtggcagaagctgtgggtgtgtgttcaaaaacttctgctgttcttctac 2940
|||||
Db 2922 aacagcaacgggtggcagaagctgtgggtgtgtgttcaaaaacttctgctgttcttctac 2981

QY 2941 aaatcacaccaggacaatcatcctcccttgccagcctgctctgctcggtactcgtctacc 3000
|||||
Db 2982 aaatcacaccaggacaatcatcctcccttgccagcctgctctgctcggtactcgtctacc 3041
QY 3001 atccctctgagtcgagaaacatccagaaagactacgtgttcaagctgcacttcaagtc 3060
|||||
Db 3042 atccctctgagtcgagaaacatccagaaagactacgtgttcaagctgcacttcaagtc 3101
QY 3061 cacgtctactcttcagggcgaaagcaggtacacgtttcgaaaagtgatggaagtgc 3120
|||||
Db 3102 cacgtctactcttcagggcgaaagcaggtacacgtttcgaaaagtgatggaagtgc 3161
QY 3121 cgcagtgccaccagctctgctcgcgacccacgtgttgagccacaagaagctctctgtg 3180
|||||
Db 3162 cgcagtgccaccagctctgctcgcgacccacgtgttgagtcacaagaagctctctgtg 3221
QY 3181 tattgatggcgagcacactcgtttcccgagtggtcgtttctctggaagacgtttccctt 3240
|||||
Db 3222 tattgatggcgagcacactcgtttcccgagtggtcgtttctctggaagacgtttccctt 3281
QY 3241 ctctgttattaatgaagcctgtgtaaaattaacacactgtctgaaaaatcaaaaacatggctt 3300
|||||
Db 3282 ctctgttattaatgaagcctgtgtaaaattaacacactgtctgaaaaatcaaaaacatggctt 3341
QY 3301 cccagcagctctctctgttccacagccgctgtttttaaccccgacactctcagcgtttgaa 3360
|||||
Db 3342 cccagcagctctctctgttccacagccgctgtttttaaccccgacactctcagcgtttgaa 3401
QY 3361 tgaacagcgtctccacactccagtcctgcacgcgtggcgctgttctttagctagt 3420
|||||
Db 3402 tgaacagcgtctccacactccagtcctgcacgcgtggcgctgttctttagctagt 3461
QY 3421 ccagttattaaaacattgtcatt 3442
|||||
Db 3462 ccagttattaaaacattgtcatt 3483

RESULT 3
AAC98992
ID AAC98992 standard; cDNA; 3094 BP.
XX
AC AAC98992;
XX
XX 09-MAR-2001 (first entry)
DT Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.
DE
XX
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
OS Homo sapiens.
XX
XX WO200055320-A1.
FN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05989.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI; 2000-579444/54.
DR
XX
XX P-PSDB; AAB54227.
PT New nucleic acid that is a pancreatic cancer antigen for preventing,

PT treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
PS Claim 1; Page 664-665; 1379pp; English.
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;

Query Match 49.1%; Score 1688.6; DB 21; Length 3094;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1699; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1740 gtggtttcagacagtcagaaaggagcgccatgcggaagcactgaaaagtctcat 1799
DB 62 gtggtttcagacagtcagaaaggagcgccatgcggaagcactgaaaagtctcat 121
QY 1800 attcccgaaatttgaaccttgcacaaatttctactaattttctcaagaaattgagca 1859
DB 122 attcccgaaatttgaaccttgcacaaatttctactaattttctcaagaaattgagca 181
QY 1860 acgacttgcctgtggaaagcgccgctcaaatgccaaatcagagattaccaaagaatcgg 1919
DB 182 acgacttgcctgtggaaagcgccgctcaaatgccaaatcagagattaccaaagaatcgg 241
QY 1920 cgatgtcatgtgaagaacattcaggggcatgaagcactggcggtcacactgtggaagca 1979
DB 242 cgatgtcatgtgaagaacattcaggggcatgaagcactggcggtcacactgtggaagca 301
QY 1980 cagcagggccttgaggccctgagaatgaatcaagagctcccgcggtgagaactt 2039
DB 302 cagcagggccttgaggccctgagaatgaatcaagagctcccgcggtgagaactt 361
QY 2040 ctgcagagactttgagctgcagaaggtgtgttaccctacgcgtcaacacacttccctcg 2099
DB 362 ctgcagagactttgagctgcagaaggtgtgttaccctacgcgtcaacacacttccctcg 421
QY 2100 gccactgacccggtcattgcactacaagaggttcttggagcggtgtgcaaacacacccc 2159
DB 422 gccactgacccggtcattgcactacaagaggttcttggagcggtgtgcaaacacacccc 481
QY 2160 gccagccacgcgaacttaaggactgcgagccggttttggcagagatacagagatggt 2219
DB 482 gccagccacgcgaacttaaggactgcgagccggttttggcagagatacagagatggt 541
QY 2220 gccacagctccacggttacgatgatcaagatggagaatttccagaagctgcacgaactcaa 2279
DB 542 gccacagctccacggttacgatgatcaagatggagaatttccagaagctgcacgaactcaa 601
QY 2280 gaaagatttgcattgacaaatttggttccgggaaaggagttcatcctcgctggg 2339
DB 1740

DB 602 gaaagatttgcattgacaaatttggttccgggaaaggagttcatcctcgctggg 661
QY 2340 cagcctcagaacagctctcggggaagggtcccagcagcagcatgttctctgttcaacga 2399
DB 662 cagcctcagaacagctctcggggaagggtcccagcagcagcatgttctctgttcaacga 721
QY 2400 cgtcctgtctatacacagcagccgggggtgcggtccctccaatcagttttaaagtcacagggca 2459
DB 722 cgtcctgtctatacacagcagccgggggtgcggtccctccaatcagttttaaagtcacagggca 781
QY 2460 gctccgctctatggtcatgacgattgagagagcagagcagtggtgggggtgcccaactg 2519
DB 782 gctccgctctatggtcatgacgattgagagagcagagcagtggtgggggtgcccaactg 841
QY 2520 cctgacccctccggggccagcagcagtcctcatcgttgcgcgcagcttctcgttccagat 2579
DB 842 cctgacccctccggggccagcagcagtcctcatcgttgcgcgcagcttctcgttccagat 901
QY 2580 ggagaagtgtgttgcagcagcagatggtccattgacctggcggaagagcagcagccc 2639
DB 902 ggagaagtgtgttgcagcagcagatggtccattgacctggcggaagagcagcagccc 961
QY 2640 cgccctgagttcctgcccagcagcccccctgacaaacagtcctcctgatgaagccacccg 2699
DB 962 cgccctgagttcctgcccagcagcccccctgacaaacagtcctcctgatgaagccacccg 1021
QY 2700 ggtgaccaggaatcagagatgacctgagcgcctcgccacatcgtgagcagtcgagcagccg 2759
DB 1022 ggtgaccaggaatcagagatgacctgagcgcctcgccacatcgtgagcagtcgagcagccg 1080
QY 2760 ccgcgaccgcccgaacacaaatggtgcagctgtgtcgtggcaccgcaacacagcagcttccat 2819
DB 1081 ccgcgaccgcccgaacacaaatggtgcagctgtgtcgtggcaccgcaacacagcagcttccat 1140
QY 2820 ggtgacttgcagcagtcagtgaggaatcagttgtctggaacactgctgaggaatctcaa 2879
DB 1141 ggtgacttgcagcagtcagtgaggaatcagttgtctggaacactgctgaggaatctcaa 1200
QY 2880 aacacgacagcgtgtgcagaagcgtgggtgtgtgttcaacaaacttgcctgtcttcta 2939
DB 1201 aacacgacagcgtgtgcagaagcgtgggtgtgtgttcaacaaacttgcctgtcttcta 1260
QY 2940 caaatcacacagcagcaaatcctccttgcacactgctcgtcgtcgtcgtcgtcgtcgtcgtc 2999
DB 1261 caaatcacacagcagcaaatcctccttgcacactgctcgtcgtcgtcgtcgtcgtcgtc 1320
QY 3000 catcctccttgatgcggaacatccagaaagactcgtgttcaagctgcacttcaagtc 3059
DB 1321 catcctccttgatgcggaacatccagaaagactcgtgttcaagctgcacttcaagtc 1380
QY 3060 ccacgttactactcagggcggaagcagtacacgttgcgaaggtggaagtgat 3119
DB 1381 ccacgttactactcagggcggaagcagtacacgttgcgaaggtggaagtgat 1440
QY 3120 ccgagtgccacagcctcgtcctgcgacccacgtgtgtgagccacaagagtccttgt 3179
DB 1441 ccgagtgccacagcctcgtcctgcgacccacgtgtgtgagccacaagagtccttgt 1500
QY 3180 gtattgatggcggacacactcgtttccgcagtggtcgttcttctggaagacgttccct 3239
DB 1501 gtattgatggcggacacactcgtttccgcagtggtcgttcttctggaagacgttccct 1560
QY 3240 tcttctgtattgaagcctgttaaatcaacacgttctgaaatacaaaaacatggct 3299
DB 1561 tcttctgtattgaagcctgttaaatcaacacgttctgaaatacaaaaacatggct 1620
QY 3300 tccacgagcctcctcgttctccacagccggttttttaaccccgacctctcagcgtttga 3359
DB 1621 tccacgagcctcctcgttctccacagccggttttttaaccccgacctctcagcgtttga 1680
QY 3360 atgaacagcgtccacacactcctcagtcctcgtcagcctggcggtgttcttttagctagt 3419
DB 1681 atgaacagcgtccacacactcctcagtcctcgtcagcctggcggtgttcttttagctagt 1740

AAA41261 to AAA43419 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, xenopus and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; anticulcer; osteopathic; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticulvasant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, parkinson's, Huntington's disease, stroke), coagulation

AAA0581-96 encode human cytoskeleton associated proteins 1 to 16
 (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat
 and diagnose cancer and cell proliferative, autoimmune/inflammatory,
 vesicle trafficking, neurological, cardiovascular, cell motility,
 reproductive and muscle disorders. Pharmaceutical compositions
 containing CYSKP-2 to CYSKP-16 can be used to treat or prevent disorders
 associated with decreased expression or activity of CYSKP (claimed), for
 example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
 cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,
 asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
 diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
 and trauma. CYSKP antagonists can be used to treat or prevent a
 disorder associated with increased expression or activity of CYSKP
 (claimed).

Sequence 3620 BP: 1156 A: 773 C: 906 G: 785 T: 0 other:

Query Match 4.9%; Score 169.6; DB 21; Length 3620;
Best Local Similarity 51.4%; Pred. No. 1.5e-35;
Matches 472; Conservative 0; Mismatches 434; Indels 12; Gaps 3;

Qy	175	aaatccagatgctggatgacaccacgaggagcatttgaagtccacaaagagctcctggg	234
Db	729	aaagtaccctcttagatggaccgaatacagctgtgacctggagaaacatgccaagggg	788
Qy	235	aaggtgctgctggatgcagtttgcaaccacactcaactctggaagtgactattttgc	294
Db	789	caagtgatttgacaagtggtgaacacctcaactctctggagaaagactactttgga	848
Qy	295	ctgaggtttcctgatcacaaagatcacggtggtggctcctctctctctctctctctg	354
Db	849	cttttgttcagaaagccctgacgagaaaacctgggttagatcctgcttaagaaataaag	908
Qy	355	aaacagattagaagggccaaagcacgttggttgaagtttgggtgaattctttccgcct	414
Db	909	agacaactggaacac---ttccatggctattcacttttaatgtgaagtttatactctc	965
Qy	415	gaccacacaaactccaagaagaactcacaaggtacactgttcgcgctcaggtgaagcag	474
Db	966	gatactctcaattgagctgaagatataccacagatactctctgtgcttcagctccggag	1025
Qy	475	gacttggctcaagcaggttgacgtgtaatgacaccagcgagctctcttctgatttcacac	534
Db	1026	gacattgcctctggccgcctgcctctctttgtgactcatgctctcctctggatccctac	1085
Qy	535	attgtgcaatctgagatgggggatttt-----gatgaagccttggacagagagcaactta	588
Db	1086	acctgcaggtggaacttggtagctatgaccagaagaacatggcagcatcgacctcagt	1145
Qy	589	gcaaaaataatacatactcctcagcaagcgcactagagagcaaaaaatcgttgaatttcaac	648
Db	1146	gaattccagtttgcctctactcagactaagagctggaagaaaggtggcgagactgcac	1205
Qy	649	cataaccacattggacaaacacacgagaaatccagatttccagctccttagagatgccccgt	708
Db	1206	aaaacccacagggcttatcgccagcaacagctgattccccagtcttagaaaaatgcaaaag	1265
Qy	709	cggctagagatgtagtgaatccgggtgcaccgccggccaaaggacaggaaggcacgaagatc	768
Db	1266	aggctttccattgatgggttgactcatcatatgccaaaggactcagaaggtgtggacatc	1325
Qy	769	aacttggcgttgccacacagggaaattctagtttccagggtttcactaaatcaatgcc	828
Db	1326	agctggtggctgtgtgctaattggactctctcaatttcaaaaagacagactgcgaatcaatcgt	1385
Qy	829	ttcaacttgggcccagggtgcggaaagctgacttccaaaggaagcgctttctctcaagaatc	888
Db	1386	tttgttggccgaaaaacttaaaaatttctataaaacgcagtaacttctacataaagtc	1445
Qy	889	cggccagatgccaatagtgogtaccaggaataccttggaaattcctgatggccagtcgggat	948
Db	1446	agaccggcagagcttggaacagatttgadagtaacctatttgatttcaacttgcacaaacacccg	1505

Qy	949	ttctgcagtccttcttggaataatctgtgtagaatcatcgctctttagactttttgaa
Db	1506	gcagcgaagaactatggaaagtgtgcgtggagcatcatactttctacaggcttgtt---
Qy	1009	gagcccaaaccaagaaccgaagcccgctcctcttttagccgggagtcattcggattcagt
Db	1563	ttctccagagcagccacaagaaccttcctgaccttgggtgccaaattcgcgtatagt
Qy	1069	ggctcggaactcagaagcacg:1086
Db	1623	ggccgacccaagcacag:1640
RESULT 6		
AAH98023		
ID	AAH98023 standard; DNA; 544 BP.	
XX		
AC	AAH98023;	
XT		
DT	10-OCT-2001 (first entry)	
XX		
DE	Murine 7-transmembrane G-protein coupled receptor coding sequence #367	
XX		
KW	Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;	
KW	7-transmembrane G-protein coupled protein receptor; ds.	
OS	Mus sp.	
XX		
PN	WO200160999-A1.	
XX		
PD	23-AUG-2001.	
XX		
PF	14-FEB-2001; 2001WO-US04700.	
XX		
PR	14-FEB-2000; 2000US-0182377.	
XX		
PA	(IMCL-) IMCLONE SYSTEMS INC.	
PA	(UYPR-) UNIV PRINCETON.	
XX		
PI	Lemischka IR, Witte L, Pereira DS;	
DR	WPI; 2001-522596/57.	
XX		
PT	DNA Sequences encoding 7-transmembrane G-protein coupled protein	
PT	receptors characteristic of hematopoietic stem cells, useful for	
PT	treating leukemia -	
XX		
PS	Claim 1; Page 122; 176pp; English.	
XX		
CC	The present invention relates to murine coding sequences for	
CC	7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The	
CC	present sequence is one such murine 7TM-GPCR coding sequence. The pres-	
CC	ent sequence was derived from stromal stem cells. The present sequence	
CC	and its corresponding protein are useful in the prevention, diagnosis,	
CC	treatment of diseases associated with inappropriate 7TM-GPCR expression	
CC	7TM-GPCRs identify specific signalling molecules, to activate an	
CC	effector-signalling cascade that triggers an intracellular response and	
CC	eventually a biological effect.	
S0	Sequence 544 BP: 143 A; 131 C; 150 G; 120 T; 0 other:	

	Query Match	4.4%;	Score 151.6;	DB 22;	Length 544;
	Best Local Similarity	56.3%;	Pred. No. 4.2e-31;		
	Matches 304;	Conservative 0;	Mismatches 234;	Indels 2;	Gaps 1;
y	2279	agaaagatttgatggcattgacaactctgtggttcagggaaggagtgcacccgtctgg	2338		
b					
b	4	agcgagacctgtctggttagaaaaacctcatgctccctggggaggagtttaccgatgagg	63		
y	2339	cgagcctcagaagctctctcggggaaggggctccacgacgcatgttcttccttgtttaacy	2398		

CC neurological disorders.

XX

Sequence 1156 BP; 381 A; 202 C; 255 G; 318 T; 0 other;

Query Match 2.0%; Score 69.4; DB 22; Length 1156;
Best Local Similarity 50.1%; Pred. No. 1.4e-08;
Matches 262; Conservative 0; Mismatches 246; Indels 15; Gaps 3;

[illegible]

Search completed: December 6, 2001, 12:00:11
Job time: 11323 sec

Matches 218; Conservative 0; Mismatches 247; Indels 3; Gaps 1;

```
QY 178 atccagatgctgatgacacccaggagcatttgaagttccacaaagagctcctgggaag 237
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 411 ATCAACTGCTTATAACAGATTGTTGGAGTTTACCCCTGTCCTGGAGACACTGGCCAG 470
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 238 gtgctgctgatgcagtttcaaccactcaactcctcgtggaagtgactatttggcctc 297
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 471 GAAAGCTCGAGCCCTGGCCAGAGCTGGAGCTGCGGAGGTCACTTACTTCAGCCTC 530
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 298 gagtttctgatcacaaaaagatacaogtggctggtgatctcttaaaacccattgtgaaa 357
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 531 TGCTACTACAAACAAGCAAAATCAAGCGCGGTGGGTAGATTGGAAAAAACCTTTGAAGAAG 590
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 358 caaat---tagaaggccaagcagctgtgttgaagtgtgtgtaaatcttccgcct 414
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 591 CAGCTGGATAAATATGATTTGGAACCTACCGTCTATTTTGGAGTGGTGTATGTCCT 650
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 415 gaccacacaaactccaagaagaactcacaaaggtacctgttcgctgcaggtgaaagcag 474
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 651 TCAGTTTCTCAGCTGCAGCAGAGATTTACCAGGTATCAGTATTATCTGCAACTGAAGAAA 710
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 475 gacttggctcaagcaggttgtaogtgaatgacacagcagctctcttctgtatttcacac 534
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 711 GATATCTTGAAGGAAGTATTCCTTGTACCTTAGAACAGCAATTCAGCTAGCAGGCTTA 770
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 535 attgtcaactgatagattgggatttggatgaagccttggacagagagcacttagcaaaa 594
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 771 GCTGTTCAACGGATTTTGTGTGATTTGATGATGATGATGATGATGATGATGATGATGAT 830
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 595 aataaaatcacatcctcagcagcagcagcagcagcagcagcagcagcagcagcagcag 642
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 831 TTTGCCCTGTCTCTGTGGGATGTTTACAGATGAAAGATATGGAA 878
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
```

RESULT 2

```
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
```

```
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14
```

Query Match 1.6%; Score 53.8; DB 1; Length 7218;
Best Local Similarity 4.7%; Pred. No. 0.00023;

Matches 19; Conservative 222; Mismatches 164; Indels 0; Gaps 0;

```
QY 1058 ttccgttcagtgctcggaactcagaagcaggtctcgcactatgttaaagaaggagacata 1117
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1437 TACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1378
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 1118 agaagtgagtttgaagaagcacagcaagattctatccggagccttgcttcac 1177
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1377 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1318
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 1178 agctacagaaactcgaagtgcgtgagcagctcagcagcagcagcagcagccttaccat 1237
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1317 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1258
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 1238 ttggaagtgccgaatctccagggggccagagctgcgcggcggaggaagaaaccgaag 1297
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1257 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1198
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 1298 ttccgcgggggagcgggtgcacacccgagcctgcgcgagagagagcccgcggtta 1357
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1197 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1138
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 1358 acaagcaggcgagcggagccgctcgcgcgccacggaggaagagagaggttcgttaag 1417
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1137 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1078
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
QY 1418 ataggaccagcagagtaaacctcagcccccgagcccaagcacag 1462
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
Db 1077 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1033
      |||||..|||||..|||||..|||||..|||||..|||||..|||||..|||||..
```

RESULT 3

```
US-08-930-001-1
; Sequence 1, Application US/08930001A
; Patent No. 6281412
; GENERAL INFORMATION:
; APPLICANT: MURATA, No. 628141210
; TITLE OF INVENTION: METHOD FOR PRODUCING OSMOTOLERANT PLANTS
; FILE REFERENCE: 0230-118P
; CURRENT APPLICATION NUMBER: US/08/930,001A
; EARLIER FILING DATE: 1997-09-26
; EARLIER APPLICATION NUMBER: JAPAN 106819/1995
; EARLIER FILING DATE: 1995-03-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2400
; TYPE: DNA
; ORGANISM: Arthrobacter globiformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (361)..(2001)
; US-08-930-001-1
```

Query Match 1.4%; Score 47.4; DB 4; Length 2400;
Best Local Similarity 43.3%; Pred. No. 0.0068;

Matches 222; Conservative 0; Mismatches 291; Indels 0; Gaps 0;


```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,401
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Lisa L.
; REGISTRATION NUMBER: 38,978
; REFERENCE/DOCKET NUMBER: PER2159POO300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-852-401-1

Query Match 1.28; Score 40.8; DB 2; Length 4776;
Best Local Similarity 43.96; Pred. No. 0.56;
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 2012 tcaagagctccgcgcgcgtggagaactcttcagagacatttgagctgcagaaagtggtgttt 2071
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 797 TCAACCGCGCGCGGGGCATCAAGGAGGACACCCGCCACGCGCTGCTGGCCCGCTCAACG 856

QY 2072 acctaccgctcaacacacttctctctgcggccactgcacgcgctcatgcactacaagcagg 2131
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 857 AGATGCCGCACACCGCGCGCGCATAGTGCGCTGCGCGCGCACCGGGGTATCGCCCTGC 916

QY 2132 tcctggagcggtctgcaaacaccaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 2191
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 917 TGGTGGCGGAGCTGTCCAAACCGGTGTTCGCCGGCTTCGCCGAGGCCCTGGAGGCGCGCG 976

QY 2192 ccgctttggcagagatacagagatggtggcacagctccacgggtacgatcatcaaatggtg 2251
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 977 CGGTGGCGCGGGGTACGCCCTGCTGTGTCAACACACCCGCGTCCGGGATGACCGAGAGG 1036

QY 2252 agaatttcagaagctgcacgaactcaagaagatttgattggcattgcacaatctgtgg 2311
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1037 ACTAGTCCGATGCTATCCCGCGCGGTGGAGGGCATGGTCTTCGTGCGCCGGAGA 1096

QY 2312 ttccgggaaggagttcatccgtctgggcagccttcagcaagctctcgggggaagggtctcc 2371
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1097 TCGCAACACCGAGGCGGAGCAGCGGATCAGCCCGCAGCTACTACGAGAAGCTGCTGGCGG 1156

QY 2372 agcagcagatgttcttctgttcaacagctctcgc 2407
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1157 ACGCGGTGCGCATGGTCTTCGTCAACGCGCGCGCGC 1192

RESULT 9
US-08-949-155-49
; Sequence 49, Application US/08949155
; Patent No. 6271436
; GENERAL INFORMATION:
; APPLICANT: Piedrahita, Jorge A
; APPLICANT: Bazer, Fuller W
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Generation of Transgenic Animal Species
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE AND DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX

```


Result No.	Query	Score	Query		DB	ID	Description
			Match	Length			
C	1	862	25.0	930	10	AL566821	AL566821 AL566821
	2	784.8	22.8	1122	10	BE746268	BE746268 601579756
	3	761.2	22.1	804	11	BG747792	BG747792 602705293
	4	757.2	22.0	865	10	BE260677	BE260677 601154188
	5	755.8	22.0	784	11	BG763918	BG763918 602736888
	6	745	21.6	861	11	BG764061	BG764061 602373071
	7	726.2	21.1	901	11	BG755554	BG755554 602491544
	8	716	20.8	736	11	BF793662	BF793662 602254351
	9	705.4	20.5	898	10	BE745887	BE745887 601573513
	10	704.6	20.5	835	11	BG829192	BG829192 602753246
	11	700.2	20.3	902	11	BE910036	BE910036 601498085
	12	687.6	20.0	721	10	AU132546	AU132546 601498085

..Qv 3261 qqtaaaattaacacctgtctgaaaaatcaaaaa 3292

Db 241 T

|||||TAATAATTAACACCTGCTGAAATCAAAA 1

BE746268 1122 bp mRNA EST 15-SEP-2000
601579750F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3928572 5',
mRNA sequence.
BE746268
BE746268.1 GI:10160260
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1122)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM759 row: p column: 13
High quality sequence stop: 780.
Location/Qualifiers
1. 1122
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3928572"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pORB7; site:1: xhoI; site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
266 a 317 C 349 g 190 t

Ch 22.8%; Score 784.8; DB 10; Length 1122;
1 Similarity 94.4%; Pred. No. 6.e-187;
991; Conservative 0; Mismatches 42; Indels 11; Gaps 7;
cagagatcacgagatggtgacacagctccacgtagcatgatacagatggagaattt 2258
cagagatcacgagatggtgacacagctccacgtagcatgatacagatggagaattt 60
cagaagctgcacgaactcaagaagatttgattggcattgacaatctgtggtccggg 2318
cagaagctgcacgaactcaagaagatttgattggcattgacatctgtggtccggg 120
aggaggttcatccgtctgggcagctccagcaagctctcgggggaagggtccacgacgcg 2378
aggaggttcatccgtctgggcagctccagcaagctctcgggggaagggtccacgacgcg 180
tgttctctgttcaacgacgtctctgtatcacgacgacgggggtcgacgctccaa 2438
tgttctctgttcaacgacgtctctgtatcacgacgacgggggtcgacgctccaa 240
tgttctctgttcaacgacgtctctgtatcacgacgacgggggtcgacgctccaa 2498
tgttctctgttcaacgacgtctctgtatcacgacgacgggggtcgacgctccaa 300
tgttctctgttcaacgacgtctctgtatcacgacgacgggggtcgacgctccaa 300

```
QY 2499 cgaagtgagggtgcccccaactgctgacccctccgggcccagggcagtcacatcatcctgccc 2558
Db 301 CGAGTGGGGGGTCCCCCACTGCTGACCTCCGGGGCCAGCGCAGTCCATCATCTGCGC 360
QY 2559 cgcagcttctcggtccgagatgagagtggtgtgagagcatccagatggcattgacct 2618
Db 361 CGCCAGTCTTCGGTCGAGATGGAGAGTGGGTGTGAGGACATCCAGATGGCCATTGACCT 420
QY 2619 ggcggagaagcagcagcccgccctcctgagttcctggccagcagcccccctgacacaa 2678
Db 421 GCGGAGAGAGAGCAGCAGCCCGCCCTGAGTTCCTGGCCAGCAGCCCCCTGACAACAA 480
QY 2679 gtccctgatgaagcaccgcggctgaccagagtgatcagaggatgacctgagcgcctcgcg 2738
Db 481 GTCCCTGTATGAAGCACCAGCGGCTGACCAGGAGTCAGAGGATGACCTGAGCGCCTCGCG 540
QY 2739 cacatcgctggagcgcagcccgcccgccagcgcgcaacacaaatggtcacgctgct-ggc 2797
Db 541 CACATCGCTGGAGCGCCAGCGCCCGCACCAGCGGCAACACAAATGGTGCACGTGTGCTGGGC 600
QY 2798 accgcaaccagcgtctccatgg-tggacttcagcatcgcatgagaggaatcagttgtct 2856
Db 601 ACCGCAACACCAGCGTCTCCATGTTGGACTTCAGCATCGCAGTGCAGAAATCAGTTGTCT 660
QY 2857 ggaacctctgagggaattcaaaaacagcaacgggtggcagaagctgt-gggtggtgtt 2915
Db 661 GGAACCTCTGAGGAATTCAAAAACAGCAACGGGTGGCAGAAAGCTGTGGGTGGTGT 720
QY 2916 ccaaaactctcctgttcttctacaaatcacacaggaacaaatcctccctgtcagcct 2975
Db 721 CAAAACTTCTGCCTGTCTTCTACAAATCACACAGGAC-ATCATCCCTCTGCCAG--- 776
QY 2976 gcctctgctgcgtactcgtcaccatccctctgagtcgagaaacatccagaagacta 3035
Db 777 --CTGCTCTGCTGTAATCGTCACATACCTCTGAGTCCGAGAAATCCAGAAAGACTA 834
QY 3036 cgt-gtcaagtcgactcaagtcacagtcctactactactactcagggcggaagcagata 3094
Db 835 GGTGTGTTCAAGCGGACTTCTATGCTCTCAGGTCTACTTACTTCAAGG-GGAAAGCGAGACC 893
QY 3095 cgttcgaaaggtgagtggaagtgatccgcagtcgacccagctct 3138
Db 894 CGTCCGACAGCGGTGGAATAGTACCCCGAGGCAACAGCGCT 937

RESULT 3
BG747792
LOCUS 804 bp mRNA EST 15-MAY-2001
DEFINITION 602705293F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4841975 5',
mRNA sequence.
ACCESSION BG747792
VERSION BG747792.1 GI:14058445
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1675 row: j column: 24
High quality sequence stop: 796.
```

```
FEATURES
Source
Location/Qualifiers
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4841975"
/tissue_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
NOTE="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library. |"
BASE COUNT 225 a 191 c 209 g 179 t
ORIGIN
Query Match 22.1%; Score 761.2; DB 11; Length 804;
Best Local Similarity 99.1%; Pred. No. 5.3e-181;
Matches 797; Conservative 0; Mismatches 3; Indels 4; Gaps 3;
QY 487 ggcaggttgacgtgtaatagcaccagcgagctctctgtattcacacattgtgcaatct 546
Db 2 GGCAGGTTGACGTGTAATGACACCGCAGCGACTCTCTTGATTTTCACACATTTGCAATCT 61
QY 547 ggaattgggatttgaagccttggacagagacacttagcaaaaataaatcata 606
Db 62 GAGATTGGGATTTTGTATGAAGCTTGGACAGAGACACTTAGCAAAAAATAATACATA 121
QY 607 cctcagcaagacgcactagaggacaaaatcgtggaatttcaccataaccacattggacaa 666
Db 122 CCTCAGCAAGACGCACCTAGAGGACAAAATCGTGGAAATTCACCATTAACCATTTGGACAA 181
QY 667 acaccgagaaatcagatttccagctcctagagattgcccgtcggctagagatgtagga 726
Db 182 ACACCAAGCAGAAATCAGATTTCCAGCTCCTAGAGATTGCCGTCGGTAGAGATGTATGA 241
QY 727 atccggttcacccgcccaggaagacaggaagcagcaagaatcaatctggcgttgcaac 786
Db 242 ATCCGTTGCAACCCGCCAAGGACAGGGAAGCGCAGAGATCAATCTGGCCGTGGCAAC 301
QY 787 acgggaattctagtttccagggtttcactaagatacaatgccttcaactgggcccagggtg 846
Db 302 ACGGAATTTCTAGTGTTCAGGTTTCAAGATCAATGCTTCACTAAGATCAATGCTTCACTG 361
QY 847 cggaaactgagcttcaagaggaagcgttctctcacaagctccggccagatgcaaatagt 906
Db 362 CGGAAGCTGAGCTTCAAGAGGAAGCGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGT 421
QY 907 ggcaccaggaatcccttgaattccctgtagtggcagtcgggattctgcaagctcctctg 966
Db 422 GCGTACAGGATACCTTGGAAATTCCTGTATGCGCCAGTCGGGATTTCTGCAAGTCTCTTGG 481
QY 967 aaaaactgtgtgaacatcatgccttctttagacttttgaagagcccaacccaaagccc 1026
Db 482 AAAATCTGTGTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAACCAAGGCC 541
QY 1027 aagccgctctctttagccgggggtcatcttccggttcagttcaggtcggactcagaagcag 1086
Db 542 AAGCCCGTCTCTTTAGCCGGGGTTCATCAATTTTCGTTTCAGTTCGGACTCGAAGCAG 601
QY 1087 gtctcgactatgtt--aagaagaggagacataagaaggtgcagtttgaaggaagacaca 1144
Db 602 GTTCTCGACTATGTTTACACGAAGGAGGACATGAAGAAGTTCAGTTTGAAGAAGCACA 661
QY 1145 gcaagattcatctatccggagccttcttccagcagctacagaaactgaatt-cggaagt 1203
Db 662 G-AAGATTTCATTCTATCCGGGCGCTTGTCTTACACAGCCTACAGAACTGAATTCGGAAGTG 720
QY 1204 ctggagcagctctcagcagagcaccagccttacaatttggagaaggtgccgaatctccagg 1263
```

```
|||||
Db 721 CTGAGCAGTCTCAGCAGACACAGCTTACATTTGAGAAGTGCCGAATCTCCAGG 780
|||||
QY 1364 ggcagagctgccggcgagaaag 1287
|||||
Db 781 GCACAGAGTCCCGCGGAGGAAG 804
|||||

RESULT 4
BE260677 865 bp mRNA EST 26-OCT-2000
LOCUS 601154188F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3510381 5',
DEFINITION mRNA sequence.
ACCESSION BE260677
VERSION BE260677.1 GI:9132166
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 865)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI94 row: o column: 22
High quality sequence stop: 730.
FEATURES
source
1..865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3510381"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 207 a 236 c 255 g 167 t
ORIGIN

Query Match 22.0%; Score 757.2; DB 10; Length 865;
Best Local Similarity 97.3%; Pred. No. 5.5e-180;
Matches 823; Conservative 0; Mismatches 18; Indels 5; Gaps 5;

QY 719 tgatgaattccggtgtgacccggccagagagaggaagcagagatcaattctggccg 778
|||||
Db 1 TGTATGAATCCGGTTGACCCGGCCCAAGGACAGGCAAGCATCAATCTGGCCG 60
|||||
QY 779 ttgccacacgggaattctagtgttcagggtttcactaagatcaatgcctcaactggg 838
|||||
Db 61 TTGCCACACGGGAATCTAGTGTTCAGGGTTTCACTAGATCAATGCCCTTCAACTGGG 120
|||||
QY 839 ccaaggtgcggaagctgagcttcaagaggaagcgctttctcatcaagctccggccagatg 898
|||||
Db 121 CCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGCCAGATG 180
|||||
QY 899 ccaatagtcgtaccaggatacattggaattctctgatgagcagtcgggattctgcaagt 958
|||||
Db 181 CCAATAGTCGTACCAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTCTGCAAGT 240
|||||
```

```
QY 959 ccttctgaaaaatctgtgttgaaacatcatgcctctcttagactttttgaagagcctcaaac 1018
|||||
Db 241 CCTTCTGGAANAATCTGTGTTGAACATCATGCCCTCTTTAGACTTTTGAAGAGCCCAAC 300
|||||
QY 1019 caaagcccaagcccgctctcttagccgggggttcattcgttcagtggtcgagactc 1078
|||||
Db 301 CAAAGCCCAAGCCGCTCTCTTTAGCCGGGGTTCATCATTTCCGTTTCAGTGGTGGACTC 360
|||||
QY 1079 agaagcaggttctcgactatgttaaaagaagagagacataagaaggttcagttgaaagga 1138
|||||
Db 361 AGAAGCAGGTTCTCGACTATGTTAAAGAGAGGAGGACATAAGAGGTTCAGTTTGAAGAAG 420
|||||
QY 1139 agcacagaagaattcattctatccggagccttgcttcacagcctcacagaactgaattcgg 1198
|||||
Db 421 AGCACAGCAAGATTCTTCTATCCGGAGCCTTGCTTCACAGCCTACAGAACTGAATTCGG 480
|||||
QY 1199 aagtgtggagcagctctcagcagagcaccagccttacattggagaaggtgcgaatctc 1258
|||||
Db 481 AAGTGCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGGTGCCGAATCTC 540
|||||
QY 1259 cagggggccaagagctgccggcgaggaagaaagcaccgaaggtttccgcggggagccggggt 1318
|||||
Db 541 CAGGGGCCAGAGTGCCTGGCGAGGAAAGAACGAAAGGTTTCGCCGGGGAGCCGGGGT 600
|||||
QY 1319 cgaccgcagccctgcgcgagagagaagccccgcgggttaacaagcagcgagcgagcgcg 1378
|||||
Db 601 CGCACCCGAGCCCTGCGCCGAGGAGAGAGCCCGGGGTAAACAAGCAGCGGAGCGAGCCG 660
|||||
QY 1379 cctcggcgccacggaggaagagagaggtcgttaagtagagaccagcagagtagtaaac 1438
|||||
Db 661 CCTCGGCGCCACGGAGAGAGAGAGGAGGTGCG-TAAGGATAGGA-CCAGCAGAGTGAAC 718
|||||
QY 1439 ctcagccccgcagcacaagcacaggtccctgact-ggcagctcacccttcagactg 1497
|||||
Db 719 CTCAGCCCCGCGAG-CAAGCACAGGCTCCTGACTGGCGAGTCTCCTACCTTTCCGAGCTG 777
|||||
QY 1498 tctgtgaactcgcagggggagtgcccttgcacaaagtgcacctgtctcccaacctgagc 1557
|||||
Db 778 TCTGTGACTCGCAGGGGAGTGGCCCTCGGCCACGTAACC-TGTTTCCCAACCTGGC 836
|||||
QY 1558 cccgac 1563
Db 837 CCGAAC 842

RESULT 5
BG763918 784 bp mRNA EST 15-MAY-2001
LOCUS 602736888F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:486244 5',
DEFINITION mRNA sequence.
ACCESSION BG763918
VERSION BG763918.1 GI:14074571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 784)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCPD/BTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI722 row: o column: 21
High quality sequence stop: 761.
```

QY	1890	tgcccaaatcagagattaccacaagaatcggcgatgtcatgc-tgaagaaacatttcagggca	1948	
Db	722	TGCCAAATCAGAGATTACCAAGAATCGCGATGTGTCATGCTTGACGAACATTTCAGGCCA	781	
QY	1949	tga 1951		
Db	782	TGA 784		
RESULT	6			
LOCUS	BG764061	861 bp	mRNA	EST 15-MAY-2001
DEFINITION	602737071F1 NIH_MGC_49	Homo sapiens	cDNA clone	IMAGE:4862417 5',
ACCESSION	BG764061			
VERSION	BG764061.1	GI:14074714		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 861)			
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue procurement: ATCC/DCTD/BTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM1722 row: n column: 18 High quality sequence stop: 817. Location/Qualifiers 1. .861 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4862417" /clone_lib="NIH_MGC_49" /tissue_type="melanotic melanoma, high MDR (cell line)" /lab_host="DH10B (phage-resistant)" /note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. "			
BASE COUNT	198 a	245 c	270 g	147 t
ORIGIN				
Query Match	21.6%; Score 745; DB 11; Length 861;			
Best Local Similarity	96.9%; Pred. No. 6.6e-177;			
Matches	791; Conservative 0; Mismatches 21; Indels 4; Gaps 3;			
QY	1910	aaagaatcggcgatgtcatgtcgtgaagaacattcagggcattcaggacacctggcgctcacc	1969	
Db	2	AAAGAATCGCGATGTCTGCTGAAGAATTCAGGCGCATGAAGACCTGGCGCTCACC	61	
QY	1970	tgtgaagcacagcgagcgcttgagcgcttgaggaatggaatcaagagctcccgccgc	2029	
Db	62	TGTGAAGCACACGAGGAGGCTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGC	121	
QY	2030	tggaacactctcagagactttgagctgcagaaggtgtgttacctaccgctcaaacct	2089	
Db	122	TGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTTGTGTACCTACCGCTCAACACT	181	

Db	722	GACAACTCTGTGGTGCCCGGGAACGGGAGTTCCATCCCGGTCTGGGCGACGTTCAAGCAAGC	781
Qy	2352	-gctctcggggaaaggggtctccagcagcgcatgttcttctctgttcaacagctcct-gcta	2409
Db	782	TCCTCGGGGAAGGACTCCCGAGCAGCGGCATGCTCTCTTGTCAACGAAGTCTCGGCTT	841
Qy	2410	tacacgagcgggggtgagcggtcccaatcagtttaaaagtcacggcgactcc	2464
Db	842	TACACGAGCGGGGTGACAGCGGTCCAAATCAGGTTAAAGCAAGGCGAGTCCC	896
RESULT 8			
LOCUS	BF793662	736 bp	mRNA
DEFINITION	602254351F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4346509		5', 12-JAN-2001
ACCESSION	BF793662		
VERSION	BF793662.1 GI:12098716		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 736)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-femail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLML at: http://image.llnl.gov Plate: L1AM968 row: f column: 14 High quality sequence stop: 710.		
FEATURES	Location/Qualifiers		
source	1..736		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4346509"		
	/clone_lib="NIH_MGC_84"		
	/tissue_type="adrenal cortex carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:1; Noti; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	167 a	228 c	209 g
ORIGIN	132 t		
Query Match	20.8%; Score 716; DB 11; Length 736;		
Best Local Similarity	99.2%; Pred. No. 1.3e-169;		
Matches	730; Conservative	0; Mismatches	5; Indels 1; Gaps 1;
Qy	2397	cgacgtcctgtatcacgagccgggggtgacggctcccaatcagtttaagtccacgg	2456
Db	1	CGAGTCCTCGTATACGAGCGGGGCTCAGCGCTCCCAATCAGTTTAAAGTCCACGG	60
Qy	2457	gcagctccgctctatggcatgacgattgagagagcgagtcagtgagggtgcctcca	2516
Db	61	GCAGTCCCGCTCATGGCATGAGCATTTGAGGAGAGCGAAGACGAGTGGGGGTGCCCA	120
Qy	2517	ctgctgacctccgggcccagcgagtcattcatcgtgcccagttctcgtccga	2576
Db	121	CTGCTGACCCCTCGGGGCCAGCGGCAGTCCATCATCTGTCGCCGCGGATCTCTCGTCCGA	180
Qy	2577	gatggagaagtgggttagggacatccagatggccattgacctgcccggagagcagcag	2636

/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 236 a 229 c 269 g 164 t
ORIGIN

Query Match 20.5%; Score 705.4; DB 10; Length 898;
Best Local Similarity 96.8%; Pred. No. 6.5e-167;
Matches 762; Conservative 0; Mismatches 21; Indels 4; Gaps 4;

QY 561 ggacaaacaccagcagaatacagatttcacagctcctagagattcccgctcggttagagatg 720
DB 1 GGACAAACACCAGCAGAAATCAGATTTCCAGCTCCTAGAGATTGCCGCTCGCTAGAGATG 60
QY 721 tatggaatccggttgaccgcgccaagacagaggaagggcacgaagatcaatctggccgtt 780
DB 61 TATGGAAATCGGTTCACCCGGCCAGGACAGGAGGACGACGAGATCAATCTGGCGGTT 120
QY 781 gccaaacacgggaattctagtgtttcagggtttcactaagaatcaatgccttcaactgggcc 840
DB 121 GCCAAACACGGGAATCTTAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCC 180
QY 841 aaggtcggaagctgagcttcaagaggaagcgtttctcctaagctccggccagatgcc 900
DB 181 AAGGTGCGGAAGTGAGCTTCAAGAGGAAGCGCTTTCTCATCAAGTCCGGCCGAGATGCC 240
QY 901 aatagtcgtaccagatcacttggaattcctgatggcagtcggagatttctcaagtcc 960
DB 241 AATAGTCCGTACCAAGATACCTTGGAAATCTCTGATGGCCAGTGGGATTTCTGCAAGTCC 300
QY 961 ttctggaaaatctgtgtgaacatctgccttcttttagactttttgaagagcccaaacca 1020
DB 301 TTCTGGAAAATCTGTG-TGAACATCATGCTCTTTTAGACTTTTGAAGAGCCCAAAACCA 359
QY 1021 aagccaagccctcctcttttagccgggggttcattcatttcggttcaggtcgagctcag 1080
DB 360 AAGCCCAAGCCCGTCCCTTTTAGCCGGGGGTCTATCAATTCGGTTCAGTGTCTCGACTCAG 419
QY 1081 aagcaggttctcgactatgttaa-agaagaggacataaagaaggtgcagtttgaaggaa 1139
DB 420 AAGCAGGTTCTCGACTATGTTAAACAGAGGAGGACACATAAGAAGTCCAGTTTGAAGGAA 479
QY 1140 gcacagcaagatttctctatccggagccttgcttcttcacagcctcacagaactgtaattcgga 1199
DB 480 GCACAGCAAGATTCTATCTATCCGGAGCCCTTGCTTACAGCCTTACAGAACTGAATTCGGA 539
QY 1200 agtgcgtgagcagctctcagcagacaccagccttaccatttgagaaggttcgcaaatctcc 1259
DB 540 AGTGTGGAGCAGTCTCAGCAGAGCACCGCCCTTACATTTGGAGAGGTCGCCGAATCTCC 599
QY 1260 agggggccagagctgcggcgaggaagaaacccgaaggtttccgcggggagccggggtc 1319
DB 600 A-GGGGCCAGAGTCCGGCGGAGAAAGGAAACGAGGTTTCCGCCGGGAGCGCGGGTC 658
QY 1320 gcaccogagccctgcggcaggagaagcccgccgggttaacaagcaggcgagcgccgc 1379
DB 659 GCACCCGAGCCCTCGCCGAGGAGAAAGCCCGGGGTAAACAAGCAGCGGACGAGCGC- 717
QY 1380 ctgcggccacagaggaagagaggaggtcgttaaggataggaccagcagagtaaac 1439
DB 718 CTGGCCGCCACCGAGGAGAAAGAGGAGGCGCGCTAAGATAGGAGACCCCAAGAGAGTAAAC 777
QY 1440 tcagccc 1446
DB 778 TTAGCCC 784

RESULT 10
BG829192
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

1 (bases 1 to 835)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue procurement: ATCC
CDNA Library preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1808 row: k column: 11
High quality sequence stop: 835.
Location/Qualifiers
1. .835

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906114"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 213 a 231 c 243 g 148 t
ORIGIN

Query Match 20.5%; Score 704.6; DB 11; Length 835;
Best Local Similarity 95.8%; Pred. No. 1e-166;
Matches 800; Conservative 0; Mismatches 24; Indels 11; Gaps 7;

QY 976 gtgaacatcatgcctcttttagacttttgaagagcccaaaccaagcccgctc 1035
DB 2 GTTGAACATCATGCCCTCTTTTA-ACITTTTGAAGAGCCCAACCAAGCCAGCCGCTC 60
QY 1036 ctctttagccgggggtcattcttcggttcagttcggttcggaactcagaagcaggttctcgac 1095
DB 61 CTCTTTAGCCGGGGTTCATCATTTTCGTTTCAGTGGTGGACTCAGAAGCAGGTTCTCGAC 120
QY 1096 tatgttaagaagaggagacataaagaaggtgcagtttgaaaggagacacaagattcat 1155
DB 121 TATGTTAAAGAAGGAGGACATAAGAAGGTGCAGTGTGAAAGGAGCACACAAAGATTCT 180
QY 1156 tctatccgagccttcttcacagcctcacagaactcagaattcgaagtcgtggagcagctct 1215
DB 181 TCATCCGGAGCCTTGCCTTCACAGCCCTACAGAACTGAATTCGGAAGTCTGGAGAGTCT 240
QY 1216 cagcagagcaccagccttaccatttgagaaggttcgcgaatctcccagggggccagagctgc 1275
DB 241 CAGCAGAGCACCAGCCTTACATTTTGGAGAAGGTGCCGAATCTCCAGGGGCCAGAGCTGC 300
QY 1276 cggcggaggaagaaacccgaaggtttccgcggggagccgggttcgcaccgcagccctcg 1335
|||||


```
Db 301 CGCCAGGAAGAACCGAAGGTTTCCCGCGGGAGCGGGTCCACCCAGCCGCTCGG 360
Qy 1336 ccgaggagaagcccccgcgggtaacaagcagcgaggagccgcctcgccgccccacgag 1395
Db 361 CCGAGGAGAAGCCCGCGGGTAACAAGCAGCGGAGCGGCGCTCGCGCCACCGAG 420
Qy 1396 gaagaggaggagtgctt-aaggtatagaccacagagagtaaaacctcagccccagacc 1454
Db 421 GAAGAGGAGGAGGTGTTCAAGGATAGACCCACAGAGTAACCTCAGCCCCCGACGC 480
Qy 1455 aagcacaggtccctgactgagcagccctcaccttccagctgtctgtgaactcgaggg 1514
Db 481 AAGCACAGGCTCCCTGACTGCGAGTCTTACCTTTCCAGAGTGTGTGTAAGTCCAGG 540
Qy 1515 gggagtgccctgcacacgtgacctgtctcccaacctgagccccgacacacagcagc 1574
Db 541 GGGAGTGGCCCTGCGCAACGTGACCTTGTCTCCAGAGACTGAGCCCCGAAACCAAGCAGC 600
Qy 1575 ctctcccttgatcagcccgctgc-tgaatgaccagcctgcccccgagcagcagatgagg 1633
Db 601 CTCTCCCTTGATCAGCCCGCTGCTTGAATGACCAAGGCTGCCCCCGGACGACGATGAG 660
Qy 1634 atgagggcggaggaagagattcccaactgataaagcgtacttca-tagctaaagaaagt 1692
Db 661 ATGAGGCGCGGAGGAGAGATTCCTCCACTGTATAAGCGTACTTCAATTAGTAAGAAAGT 720
Qy 1693 tctaccacggcgaac---atatctgaaggtatctcgaag-ttatcacttggtgttca 1748
Db 721 TCTACCACCGAGCAACCATATTCTGAAGGATCTCGAAGTTTATCACTTCGCGGTTCAG 780
Qy 1749 gagcagctgagcaagaggacgcctcccggaagcac---tgaagaagtctcata 1800
Db 781 AAGCACAGTGAAGCAAGAGGAGCGCATGCCGGAAGACACTTGAAGAAGTCTCAT 835

RESULT 11
BE910036 601498085F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3900021 5',
LOCUS mRNA sequence. EST 20-OCT-2000
DEFINITION BE910036
ACCESSION BE910036
VERSION BE910036.1 GI:10406227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 902)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM9698 row: j column: 22
High quality sequence stop: 750.
Location/Qualifiers
1. 902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3900021"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
```

```
BASE COUNT 231 a 248 c 272 g 151 t
ORIGIN
Technology."
Query Match 20.3%; Score 700.2; DB 11; Length 902;
Best Local Similarity 95.8%; Pred. No. 1.3e-165;
Matches 752; Conservative 0; Mismatches 28; Indels 5; Gaps 3;
Qy 1191 gaattcgaagtgcctggagcagctcagcagagcaccagccttacatttggagaaggtgc 1250
Db 1 GAAATTCGGAAGTGTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGGTGC 60
Qy 1251 cgaatctccagggggccagagctgcggcgaggaaagaaacccgaaggtttccgcgggga 1310
Db 61 CGAATCTCCA-GGGGCCAGAGTGCCTGGAGAAAGAAAGAAAGAGGTTCCTCGCGGGGA 119
Qy 1311 gccggggtcgcacccagcagcctgcgcgagagagagcccccgcgggtaacaagcagcagc 1370
Db 120 GCCGGGGTGCACCCGAGCCTTGCCTCCGAGGAGAGCCCGGGTAAACAGCAGCCGA 179
Qy 1371 cggagccgcctcggcgccacgagagagagagaggtcggttaaggatagaccaccagca 1430
Db 180 CGAGCCCGCTCGGCGCCACCGAGAGAGAGAGAGGTCTGTTAAGGATAGACCCAGCA 239
Qy 1431 gagtaaacctcagcccccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1490
Db 240 GAGTAACCTCAGCCCGCAGCAGCAAGCAGCAGCTCCCTGACTGGCAGTCTCAGCTTTC 299
Qy 1491 cgagctgtctgtaactgcgagggggagtgccctgccaaagcagcagcagcagcagcagc 1550
Db 300 CGAGCTGTCTGTGAATCGCAGGGGGAGTGGCCCTTGCACACGTGACCTTGTCTCCCAA 359
Qy 1551 cctgagccccacacacagcagcagcctctcccttgatcagccgcgtgtaagcagcagc 1610
Db 360 CTTGAGCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 419
Qy 1611 ctgccccggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1670
Db 420 CTGCCCCCGGAGCGAGCATGAGGATGAGGGCGGAGGAGAGATTCCTCACTGATAAAGC 479
Qy 1671 gtacttcagtaaggagagtgctaccacccagcagcagcagcagcagcagcagcagcagc 1730
Db 480 GTACTTCATAGCTAAGAGTGTCTTACCACCGAGCGAGCATATCTGAAGGATCTCGAAGT 539
Qy 1731 tatcacttcgtgttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1789
Db 540 TATCACTTCGTGGTTTCAGAGCAGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
Qy 1790 aaa---gtctcatattcccgaaattttgaacctttgcacaaatttcataattttctca 1846
Db 600 AAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAATTTTCATAATAATTTCTCA 659
Qy 1847 aggaattgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1906
Db 660 AGGAATTTGAGCAACGACTTGCCTGTGGGAAAGCCGCTCAATATGCCAAATTCAGAGATT 719
Qy 1907 accaagaatcgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1966
Db 720 ACCAAGAATCGCGATGTCTGCTGAAGAACCATTTAGCCATTAAGAACCCCGGGGGTCC 779
Qy 1967 acctg 1971
Db 780 CTGTG 784

RESULT 12
AUI32546
LOCUS AUI32546
DEFINITION AUI32546 NT2RP4 Homo sapiens cDNA clone NT2RP4000050 5', mRNA
ACCESSION AUI32546
VERSION AUI32546.1 GI:10992900
```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 721)
Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y., Saito, K., Yamamoto, J.,
Nishikawa, T., Nakamura, Y., Nagai, T., Sugano, S., Masuho, Y. and
Isogai, T.

TITLE HRI human cDNA project (Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y.,
Saito, K., Yamamoto, J., Nishikawa, T., Nakamura, Y., Nagai, T., Sugano
, S., Masuho, Y., Isogai, T.)

JOURNAL Unpublished (2000)
COMMENT Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp

FEATURES
Source HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. 721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP400050"
/cell_lib="NT2RP4"
/cell_type="teratocarcinoma"
/note="Vector: pME18SPL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 183 a 180 c 214 g 141 t 3 others

Query Match 20.0%; Score 687.6; DB 10; Length 721;
Best Local Similarity 98.8%; Pred. No. 1.9e-162;
Matches 712; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 709 cggctagatgctatggaatccggttgccaccggcccaagagcaggaagcaggaagtc 768
DB 1 CGGCTAGATGATGGAATCCGTTGCCACCCGGCCCAAGCAGGGAAGCAGGAAGATC 60

QY 769 aatctggcgttgcccaacacgggaattctagtttcagggtttcactaatgac 828
DB 61 AATCTGGCCGTTGCCAACACGGGAATCTAGTCTTTTCAGGGTTTCACTAAGATCAATGCC 120

QY 829 ttcaactggcccaagtgccggaagctgagcttcaagaggaagcgctttctcatcaagctc 888
DB 121 TTCACTGGGCCAAGTGCGGAAGCTGAGCTTCAACAGGAAGCGCTTCTCATCAAGCTC 180

QY 889 cgcccgatgccaatagtcgtaccaggtatccttggaattcctgtagtgcagtcgggat 948
DB 181 CGCCAGATGCCAATAGTCGTCACAGGATACCTTGGAAATCTGTATGTCGTCGGAT 240

QY 949 ttctgaagctctctggaaaatctgttgaaatcatcgtctctttgactttttgaa 1008
DB 241 TTCTGCAAGTCCTCTGGAATAATCTGTGTGAACATCATGCGCTCTTTTACATTTTGA 300

QY 1009 gagcccaaaccaagcccaagcccgctcttttagcgggggtcatcttgcgttcagt 1068
DB 301 GAGCCCAACCAAGCCCAAGCCCGTCTTTTAGCCGGGGTCATCATTTCCGTTCACT 360

QY 1069 ggtcggactcagaagcaggtttctcgactatgtttaaagaaggagacataagaaggtgcag 1128
DB 361 GGTCCGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAAGGAGGACATAAGAAGGTGCAG 420

QY 1129 ttgaaagaagcacagcagattctattccggagccttcttcacagcctacagaa 1188
DB 421 TTTGAAAGGAAACACAGCAAGATTCATCTATCCGGAGCCTTGTCTTCACAGCCTACGGAA 480

QY 1189 ctgaattcgaagtgctgagcagctctcagcagagcaccagccttaccatttgagaaggt 1248
DB 481 CTGAATTCGAAGTGCTGGAGCAGTCTACGAGAGCACCAGCCTTACATTTGGAGAAGGT 540

QY 1249 gccgaatctccggggggccagagctcggcgaggagaaagaaagcgaagtttcgcggg 1308
DB 541 GCGAATCTCCAGGGGGCCAGAGCTCCGCGGAGGAAGAACCGAAGGTTTCCGCCGG 600

QY 1309 gaaq-ccgaggtcgaccccgagccctcgccgagagagaagcccgcggttaacaagcagcc 1367
DB 601 GAGCCCGGGGTGCACCCGAGCCCTCGCCGAGGAAGCCCGCGGTAAACAGCAGGC 660

QY 1368 ggacggagcgcctcgcccgccacgaggaagaggagggtcg-ttaaggtatagaccc 1426
DB 661 GGAGGGAGCCGCTNGCGCCACGAGGAGGAAGANGTCTGTTAAGGATAGGACCC 720

QY 1427 a 1427
DB 721 A 721

RESULT 13
LOCUS BG323704 942 bp mRNA 27-FEB-2001
DEFINITION 60242195F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4560191 5',
mRNA sequence.
ACCESSION BG323704
VERSION BG323704.1 GI:13130141
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 942)
NIH-MGC http://mgc.mci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DIRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1268 row: m column: 24
High quality sequence stop: 760.

FEATURES
Location/Qualifiers
1. 942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4560191"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 237 a 264 c 286 g 155 t

ORIGIN
Query Match 19.7%; Score 679.6; DB 11; Length 942;
Best Local Similarity 96.8%; Pred. No. 2.1e-160;
Matches 736; Conservative 0; Mismatches 19; Indels 5; Gaps 4;

```
QY 1145 gcaagattcattctcagagccttgcctacagccctacagaaactgaatcggagtc 1204
Db 1 GCAAGATTCTATCTATCGGAGCCTTGCTTACAGCCCTACAGAACTGAATTCGGAAGTC 60
QY 1205 ttagcagctcagcagagcaccagccttacatttggagaagtgccgaatccaggg 1264
Db 61 TGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGTGCCGAATCTCA-GGG 119
QY 1265 gccagagctgcggcgaggaaaggaacgaaggtttccgcggggagcggggtcgacc 1324
Db 120 GCCAGAGTGCCTCGGAGGAGAAAGAACGAGGTTTCCGCCGGGAGCGGGTGCAC 179
QY 1325 ccagccctgcggcagagaaagcccggttaacagcagcagcagcagcgcctcg 1384
Db 180 CGAGCCCTGCCTCGGAGGAGAACCCCGGGTAAACAGCAGCGGAGCGGCTCGG 239
QY 1385 cccccacagagagagagaggtcgcttaaggaagagccagcagagagaaacctcag 1444
Db 240 CCCCCACGAGGAGAGGAGGAGGTCGTTAAGATAGGACCCAGCAGAGTAACCTCAGC 299
QY 1445 cccccagcagcagcagcagcctcctgactgagcagtcctcactcactccttcagctgtgtga 1504
Db 300 CCCCCAGCAGCAGCAGCAGCTCCTGACTGCGAGTCTCACCCTTCCGAGCTGTCTGA 359
QY 1505 actcagggggagtgccctgcacagtgacactgtacttctcccaacctgagcccgaca 1564
Db 360 ACTCAGGGGGGAGTGCCCTGCAAGCTGACCTGTCTCCCAACCTGAGCCCGGACA 419
QY 1565 ccaagcagcctcctccttgatcagccgctgctgaatgaccagcgcctgcgcccgagcg 1624
Db 420 CCAAGCAGGCTCTCCTTGATCAGCCGCTGCTGAATGACGAGG-CYGCCCCCGGACGG 478
QY 1625 acgatgagatgagggcgaggagagagattcccaactgataaagcgtacttcagctta 1684
Db 479 ACGATGAGATGAGGGCGGAGGAGAGATCCCAACTGATAAAGCGTACTTCATAGCTA 538
QY 1685 aggaagtgtaccaccagcagaaacatatctgaagatctcgaagtattcaacttcgtgtg 1744
Db 539 AGGAAGTGTCTACCCAGCGGAGGAGATCTGAAGGATCTCGAAGTATFACATTCGTGTG 598
QY 1745 ttcagagcacagtgcagaaagagcgcctcgcggaagcactgaaagtctcatattcc 1804
Db 599 TTCAGAGCACAGTGCAGAAAGAGGCGCATGCCGGAAGCACTGAAAAGTCTCATAGTCC 658
QY 1805 cgaatttgaaaccttgacaaattctactaattttctcgaagaaatttcagcagcagc 1864
Db 659 CGAATTGGAACCTTGCACAAATTTCTACTTAATTTCTCCA--GGAATTGAGCAAGGAC 716
QY 1865 ttgcctgtggaagccgcctcaaatgcccaaatcagaga 1904
Db 717 TTG-CCTGTGGGAGGCCGCTCAAAATGCCAAATCAGAGATA 755
```

```
RESULT 14
LOCUS BG767698
DEFINITION 602741690F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4871374 5',
mRNA sequence.
ACCESSION BG767698
VERSION BG767698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
```

```
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1746 row: c column: 23
High quality sequence stop: 811.
FEATURES
location/Qualifiers
source
1..891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4871374"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into ECORI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 248 a 216 c 249 g 177 t 1 others
ORIGIN
Query Match 19.7%; Score 678.2; DB 11; Length 891;
Best Local Similarity 94.2%; Pred. No. 4.7e-160;
Matches 801; Conservative 0; Mismatches 39; Indels 10; Gaps 9;
QY 484 caagcaggttgacgtgaatgacaccag-cgcagctctcttgatttcacacattgtgca 542
Db 2 CAAGCAGGTTGACGTGTAATGACACCAGCAGCAGCTCTCTTGATTTCACACA-TGTGCA 60
QY 543 atctgagattgggattttgatgaagccttgacagagagcacttagcaaaaaataa 602
Db 61 ATCTGAGAT--GGGGATTTTGATCAAGCCTTGACAGAGAGACACTTAGCAAAAAATAA 118
QY 603 catactcagcagcagcactagagacaaaatcgtggaatttcacataaccacattgg 662
Db 119 CATACTCAGCAAGACGCACTAGAGGACAAAATCGTGGAAATTCACCATACCACA-TGG 177
QY 663 acaaacacagcagaaatcagatttccagctcctagatgtgccctcggtcagagatgta 722
Db 178 ACAAAACACAGCAGAAATCAGATTTCAGCTCCTAGAGATTGCCCGTCGGCTAGAGATGA 237
QY 723 tggaaatccggttgaccgcggcccaagacaggaaggaagcagaaagatcaatctgcccgttc 782
Db 238 TGGAAATCCGGTTGCACCCCGCCCAAGGACAGGAGGACGAGATCAATCTGGCCGTTGC 297
QY 783 caacacgggaattctagtgtttcaggggtttcactaagatcaatgcctcaactgggcca 842
Db 298 CAACACGGGAATCTAGTCTTTTCAGGGTTTCACTAAGATCAATGCCCTTCAACTGGCCAA 357
QY 843 ggtgcggaagctgagcttcaagaggaagcgcttctcctcatcaagctccgcccagatgccaa 902
Db 358 GGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGCCAGATGCCAA 417
QY 903 tagtgcgtaccagataaccttggaattcctgagtcgagcagctgggattcttcaagtcctt 962
Db 418 TAGTGCCTACCAAGGATACCTTGGAAATCTGATGGCCATCGGGATTTCTGAAGTCCCT 477
QY 963 ctggaatactgtgtgaacatcatgctctttagacttttgaagaccccaaacaaa 1022
Db 478 CTGGAAAATCTGTGNGAACATCATGCTCTCTTTAGACTTTTGAAGACGCCAACCAAA 537
QY 1023 gcccaagccgctcctcttttagcgggggttcatacttcggttcagtggtcgactcagaa 1082
Db 538 GCCCAAGCCGCTCTCTTTAGCGGGGGTCAATCATTTTCGTTTCAGTGGTGGACTCAGAA 597
QY 1083 gcaggtctctcagctatggt-aaagaagggaggacataagaaggtgcagtttgaaggagac 1141
```

Db	598	GCAGGTTCTCGACATATGTAAAGAAGAGGACATAAAGAGTGCAG-TTGAAAGGAAGC	656
Qy	1142	acagcaagattcattctatccggagccttggttcacagcctcacagaactgaattcg-aa	1200
Db	657	ACACAAGATTCAATTCATCCGGAGCCTTGCTGGCACAGCTACAGAACTGAATTCGGAAA	716
Qy	1201	gtgctggagcagtctcagcagcagaccagccttacctattggagaaggtygcgaatctcca	1260
Db	717	GTGTGGAGCAGTCTCAGCAGAGCACCAGGCTTACATTTGGAGAAGTGCCGAATCTCCA	776
Qy	1261	gggggccagactccggcggaggaagaaaccgaaggtttccgcggggagccggggtcg	1320
Db	777	-GGGCCAGAGCTG-CGGCGAGGAAGAACGAAGGTTCCCGCGGAGCGGTGCACCCG	834
Qy	1321	cacccgagcc 1330	
Db	835	AGCTGC GCC 844	
RESULT	15		
BF206296			
LOCUS	BF206296	883 bp mRNA EST	06-NOV-2000
DEFINITION	60186962F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4100044 5',		
ACCESSION	BF206296		
VERSION	BF206296.1	GI:1109882	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NTH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: ATCC		
	CNA Library Preparation: Ling Hong/Rubin Laboratory		
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov		
	Plate: LLCM969 row: i column: 05		
	High quality sequence stop: 722.		
FEATURES	Location/Qualifiers		
source	1..883		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4100044"		
	/clone_lib="NIH_MGC_19"		
	/tissue_type="neuroblastoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:		
	ECORI; cDNA made by oligo-dT priming. Directionally		
	cloned into EcoRI/XhoI sites using the following 5'		
	adaptor: GGCACGAG(G). Library constructed by Ling Hong		
	in the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies).		
	Note: this is a NIH_MGC Library."		
BASE COUNT	192 a 264 c 266 g 161 t		
ORIGIN			
Query Match	19.6%	Score 673; DB 11; Length 883;	
Best Local Similarity	94.7%;	Pred. No. 9.5e-159;	
Matches	772; Conservative	0; Mismatches 35; Indels 8; Gaps 7;	
Qy	2346	cagcaagctctcggggaagggtccagcagcatgttctcttgttaaacagctcct	2405
Db	2	CAGCAAGCTCTCGGGGAAGGGGCTCCAGACGGCATGTTCTTCTGTTCACAGCTCCT	61

Qy	2406	gctatacacgagccgagggggctgacggccctccaatcaagttaaagttccacggggcagctccc	2405
Db	62	GCTATACACGAGCGGGGGTGCAGGCCCTCCAAATCAGTTTAAAGTTCACAGGCGACGTCCC	121
Qy	2466	gctctatggcatacgaattgagagagacgaacgaagtggggggtgccccactccctgac	2525
Db	122	GCTCTATGGCATGACGA -TGAGGAGAGCGAAGACAGAGTGGGGGTGCCCCACTGCGCTGAC	180
Qy	2526	ctctccgggggcagcgggcagtcctcatcgtggtggccagttctcgttcgagatggagaa	2585
Db	181	CCTCCGGGGCCAGCGGCGAGTCCATCATCTGTGGCGCCAGTTCTCGTCCGAGATGGAA	240
Qy	2586	gtgggttggagacatccagatggccattgacctggcgggaagacgacgagcccccgccc	2645
Db	241	GTGGGTTTGAGGACATCCAGATGGCCATTGACCTGGCGGAGAGACGACGAGCCCCGCC	300
Qy	2646	tgaattcctggccagcagccccctgaacaaagtcctcgatgaagccacgcggcgtga	2705
Db	301	TGAGTTCTTGCCGACGAGCCCCCTGACAAACAGTCCCTGTATGAAGCCACCGCGGCTGA	360
Qy	2706	ccaggagtcagagatgacctgagcgttcgcgcacatcgtcggagcgcagggccccgcga	2765
Db	361	CCAGGAGTCAGAGGATGACCTGAGCGCCTCGCGCACATCGCTGGAGCGCCAGGCCCGCA	420
Qy	2766	ccgcggcacaacaatggtgcactgtgctggaacgcgaacacacagcgtctccatggtgga	2825
Db	421	CCG -GGCAACACAAATGTGTGCACGTGTGCTGGCAGCCGCAACACACAGCGTCTCCATGGTGG	479
Qy	2826	cttcagatccgagtgagaatacagttctggaacctgctggaacctgctgaggaattcaaaaacag	2885
Db	480	CTTCAGCATCGAGTGAGGAATCAGTTGCTGGAACCTGCTGAGGAATTCAAAACAG	539
Qy	2886	caacggtggcagaagcgtgggtgggtgttcacaaattct -gcctgttcttcttacaagt	2944
Db	540	CAACGGGTGGCAGACGCTGGGTGCTGTTCACAACTTCCTGCGCTGTCTTCTACAAAT	599
Qy	2945	cacaccaggacaatctcccttg -ccagcctgctctgctcgtcggtactcgtcacaatc	3003
Db	600	CACACCAGGACAAATCATCCCCCTGTCCAGCGCGCCTCTGCTCGGCTTACTGCTCACATCC	659
Qy	3004	ccctctgagtcga -gaacatccagaagaactcgtttcaagctgcacttcaagtccca	3062
Db	660	CCTCTGAGTCCGAGGAACATCCAGAAAGACTACGTGTTCAAGCTGACATTCAGTGCCCA	719
Qy	3063	cg -tctactacttcaggcggaagcgaagtacacgcttcgaaaggtggatggaagtgcac	3121
Db	720	CGTTCTACTACTTCAGGCGGGAAGCGAGTACAGT -CGAAGGTGGTTGGAAGTCGATC	777
Qy	3122	gcagtgccaccagctctgctcgcgacccccagtg	3156
Db	778	GGAGTGGCCAAAGTTGCGCTCGGACCCACGTTG	812

Search completed: December 6, 2001, 10:34:44
Job time: 7016 sec

	Query Match	19.6%	Score 673;	DB 11;	Length 883;
	Best Local Similarity	94.7%	Pred. No. 9.5e-159;		
	Matches 772;	Conservative 0;	Mismatches 35;	Indels 8;	Gaps 7;
Qy	2346	cagcaagctctcggggaagggtccagcagcgcatgttcttctctgttccaacgactcct	2405		
	2				
Db	2	CAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTGTCAACGAGCTCCT	61		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:51:43 ; Search time 72.4 Seconds
(without alignments)
93.103 Million cell updates/sec

Title: US-09-555-342A-2_COPY_764_854
Perfect score: 468
Sequence: 1 GSLSKLGGKLGLOQMFLEFN.....IVAASSRSEKWKVEDIQMA 91

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	395	21 AAB54227	Human pancreatic c
2	468	100.0	1045	20 AAY07482	Human chondrocyte-
3	468	100.0	1045	21 AAY91947	Human cytoskeleton
4	152.5	32.6	432	22 AAB93568	Human protein sequ
5	122	26.1	766	21 AAB31248	Rat actin-binding
6	118.5	25.3	183	22 AAM39396	Human polypeptide
7	118.5	25.3	383	22 AAM42414	Human polypeptide
8	118.5	25.3	404	22 AAM41182	Human polypeptide
9	112.5	24.0	279	22 AAB94617	Human protein sequ
10	111.5	23.8	249	22 AAM40179	Human polypeptide
11	111.5	23.8	249	22 AAB94506	Human protein sequ

12	111.5	23.8	249	22 AAB64412	Amino acid sequenc
13	85.5	18.3	762	22 AAB94398	Human protein sequ
14	84.5	18.1	394	18 AAW26726	Human integrin reg
15	84.5	18.1	394	20 AAY26928	Human integrin reg
16	84.5	18.1	394	22 AAE00737	Human integrin reg
17	81.5	17.4	201	22 AAB93655	Human protein sequ
18	81	17.3	245	22 AAM39355	Human polypeptide
19	81	17.3	246	22 AAM41141	Human polypeptide
20	81	17.3	301	22 AAB94156	Human protein sequ
21	81	17.3	408	22 AAB95198	Human protein sequ
22	80.5	17.2	505	19 AAW61027	Murine guanine nuc
23	80.5	17.2	535	19 AAW61028	Murine guanine nuc
24	80.5	17.2	554	19 AAW61026	Murine guanine nuc
25	80.5	17.2	814	14 AAR43578	Peptide which modu
26	79.5	17.0	710	20 AAR80995	Amino acid sequenc
27	79.5	17.0	1244	21 AAY68825	Amino acid sequenc
28	77	16.5	619	22 AAB97025	Human colon carcin
29	75.5	16.1	399	18 AAW33417	Human integrin reg
30	75.5	16.1	399	20 AAY26927	Human integrin reg
31	75.5	16.1	399	22 AAE00736	Human integrin reg
32	75.5	16.1	399	22 AAE00739	Human integrin reg
33	75	16.0	417	22 AAB73491	Human cytohesin-2
34	74.5	15.9	399	22 AAE00740	Human integrin reg
35	72.5	15.5	378	22 AAM42353	Human polypeptide
36	71	15.2	129	19 AAW54285	Human modified cyt
37	71	15.2	263	18 AAW18783	Cytohesin 2 encode
38	71	15.2	398	18 AAW18782	Cytohesin 1. Homo
39	71	15.2	398	22 AAE00738	Human B2-1 protein
40	70.5	15.1	399	19 AAW57707	Mouse GRP protein
41	69.5	14.9	537	18 AAW37045	Murine alpha-(2) s
42	68	14.5	1227	22 AAB65637	Novel protein kina
43	68	14.5	1289	20 AAY27163	Peptide seq ID No:
44	68	14.5	1289	21 AAY56781	Human Trad protein
45	67.5	14.4	487	22 AAU00628	Novel human protei

ALIGNMENTS

RESULT 1
AAB54227
ID AAB54227 standard; Protein; 395 AA.
XX AC AAB54227;
XX DT 09-MAR-2001 (first entry)
XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:679.
XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytosolic; neuroprotective;
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX OS Homo sapiens.
XX PN WO200055320-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05989.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI: 2000-579444/54.
DR N-PSDB; AAC98992.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 11; Page 1115-1116; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 395 AA;

Query Match 100.0%; Score 468; DB 21; Length 395;
Best Local Similarity 100.0%; Pred. No. 8.5e-56;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSKISGKGLQRMFFLNDVLLYTSRGLTASNQPKVHGQLPLYGWTIESEDEWGVPH 60
|||||
Db 220 gslskisgkglqrmfflndvlltsrgltasnkqkvhgqlplygmtieesedewgvph 279
|||||

QY 61 CLTLRGQRQSIIVAAASRSSEMEKWKVEDIQMA 91
|||||
Db 280 cltlrgqrqsiivaaasrssemekekvediqma 310
|||||

RESULT 2
AA07482
ID AAY07482 standard; Protein; 1045 AA.
XX
AC AAY07482;
XX
DT 17-AUG-1999 (first entry)
XX
DE Human chondrocyte-derived protein CDEP.
XX
KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW Dbl1 homology domain; pleckstrin homology domain; rheumatoid arthritis;
KW drug.
XX
OS Homo sapiens.
XX
PN WO9928458-A1.
XX
XX 10-JUN-1999.
XX
XX 27-NOV-1998; 98WO-JP05348.
XX
XX 27-NOV-1997; 97JP-0342060.
XX
XX (CHUS) CHUGAI SEIYAKU KK.
XX
XX Kato Y, Kawamoto T, Koyano Y;
PI

XX WPI: 1999-371117/31.
DR N-PSDB; AAX79183.
XX
PT Protein CDEP expressed in differentiated chondrocytes, and gene
PT encoding it
XX
PS Claim 2; Fig 1; 59pp; Japanese.
XX
CC This sequence represents a protein (CDEP) expressed in differentiated
CC human foetal chondrocytes, which contains an ezrin-like domain, a Dbl
CC homology (DH) domain and a pleckstrin homology (PH) domain. The encoding
CC nucleic acid or protein can be used in the investigation and treatment of
CC cancers and arthritic diseases (including chronic rheumatoid arthritis),
CC or for screening of candidate anticancer drugs.
XX
SQ Sequence 1045 AA;

Query Match 100.0%; Score 468; DB 20; Length 1045;
Best Local Similarity 100.0%; Pred. No. 3.3e-55;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSKISGKGLQRMFFLNDVLLYTSRGLTASNQPKVHGQLPLYGWTIESEDEWGVPH 60
|||||
Db 764 gslskisgkglqrmfflndvlltsrgltasnkqkvhgqlplygmtieesedewgvph 823
|||||

QY 61 CLTLRGQRQSIIVAAASRSSEMEKWKVEDIQMA 91
|||||
Db 824 cltlrgqrqsiivaaasrssemekekvediqma 854
|||||

RESULT 3
AA91947
ID AAY91947 standard; Protein; 1045 AA.
XX
AC AAY91947;
XX
DT 19-JUL-2000 (first entry)
XX
DE Human cytoskeleton associated protein 2 (CYSKP-2).
XX
XX Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Modified-site 23
FT /note= "potential phosphorylation site"
FT Modified-site 36
FT /note= "potential phosphorylation site"
FT Modified-site 41
FT /note= "potential phosphorylation site"
FT Domain 47..85
FT /note= "signature sequence"
FT Modified-site 92
FT /note= "potential phosphorylation site"
FT Domain 94..123
FT /note= "signature sequence"
FT Domain 144..190
FT /note= "signature sequence"
FT Modified-site 150
FT /note= "potential phosphorylation site"
FT Modified-site 152
FT /note= "potential N-glycosylation site"
FT Domain 196..249
FT /note= "signature sequence"
FT Modified-site 207
FT /note= "potential phosphorylation site"
FT Domain 261..279
FT /note= "signature sequence"

